



HAL
open science

Disaggregation of Electrical Appliances using Non-Intrusive Load Monitoring

Thomas Bier

► **To cite this version:**

Thomas Bier. Disaggregation of Electrical Appliances using Non-Intrusive Load Monitoring. Other. Université de Haute Alsace - Mulhouse, 2014. English. NNT : 2014MULH8860 . tel-01314432

HAL Id: tel-01314432

<https://theses.hal.science/tel-01314432>

Submitted on 11 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2014

UNIVERSITÉ DE HAUTE-ALSACE, MULHOUSE
ÉCOLE DOCTORALE JEAN-HENRI LAMBERT
LABORATOIRE MIPS

THÈSE

présentée par

Thomas Bier

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE HAUTE-ALSACE

Discipline : Électronique, Électrotechnique et Automatique

Disaggregation of Electrical Appliances using Non-Intrusive Load Monitoring

(Arrêté Ministériel du 30 mars 1992)

Soutenue publiquement le 17 décembre 2014 devant le jury composé de:

Pr.	YASSINE RUICHEK	Université de Technologie de Belfort-Montbéliard	Rapporteur
Pr.	MICHEL PAINDAVOINE	Université de Bourgogne	Rapporteur
Pr.	ALAIN DIETERLEN	Université de Haute-Alsace	Examinateur
Pr.	DIRK BENYOUCEF	University of Furtwangen	Examinateur
Dr.	DJAFFAR OULD-ABDESLAM	Université de Haute-Alsace	Co-encadrant
Pr.	JEAN MERCKLÉ	Université de Haute-Alsace	Directeur
Pr.	PIERRE RAYMOND	European Laboratory for Sensory Intelligence	Invité

Acknowledgment

A PhD thesis cannot be performed without the help of many people. During my PhD, many people have contributed to the success of this. Especially some people be thanked here.

First, I would like to thank my supervisor, Prof. Jean Merckle for their excellent advice and constant support. His experience in the training of graduate students helped me a lot.

I acknowledge my advisor Prof. Djaffar Ould-Abdeslam from TROP for the assistance and guidance. His administrative support has simplified much.

A special thanks goes to Prof. Dirk Benyocuef for his support and motivation. Since the beginning of my studies, he has always given me constructive approaches.

I would also like to give thanks to my lab colleagues from ReSP in Germany and TROP in France, especially to my fellow student Philipp Klein.

I would like to thank my parents for their unremitting support. Since I have started to study electrical engineering they stay by my side.

Finally, I would like to acknowledge the financial support provided by the state of Baden-Wuerttemberg through the Smart Metering projects, without which my PhD would not have been possible.

Acronyms

ADALINE	ADaptive LInear NEuron
ANN	Artificial Neural Network
AWGN	Additive white Gaussian Noise
ED	Event Detection
FP	False Positives
FPR	False Positive Rate
FT	Fourier Transformation
IALM	Intrusive Appliance Load Monitoring
MLP	Multi Layer Perceptron
MS	Measurement System
MSE	Mean Squared Error
NALM	Non-Intrusive Appliance Load Monitoring
NILM	Non-Intrusive Load Monitoring
PDF	Probability Density Function
PDS	Power Density Spectrum
PPV	Positive Predicted Value
SNR	Signal to Noise Ratio
ST	S-Transformation
STFT	Short-Time Fourier Transformation
TP	True Positives
TPR	True Positive Rate

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Systems, Approaches and Applications	2
1.3	Research of the ReSP	4
1.4	Problem Statement in the Thesis	6
1.5	Structure of the Thesis	6
2	Pattern Recognition for NALM	9
2.1	General Definitions	10
2.2	Types of Classifiers	11
2.3	Evaluation of Classifiers	11
2.3.1	Confusion Matrix	12
2.3.2	Statistical Quality Criteria	12
2.3.2.1	Recall, True Positive Rate	13
2.3.2.2	Precision, Positive Predicted Value	13
2.3.2.3	Combined Measures	13
2.3.3	Determining a Rate	14
2.4	Time Frequency Analyze	14
2.4.1	Fourier Transformation	14
2.4.2	Short-Time Fourier Transformation	15
2.4.2.1	Transformation	15
2.4.2.2	Power Density Spectrum	16
2.4.3	S-Transformation	16
2.5	Conclusion	17
3	State of Art for NALM	19
3.1	State of Research	20
3.1.1	Steady State	21
3.1.2	Transient State	23
3.1.3	Newer Approaches for NALM Systems	24
3.2	State of Technology	25
3.3	Conclusion	26
4	Analyse of the Signals for NALM	27
4.1	Introduction	28
4.2	System Model	28
4.2.1	Appliance Model	28
4.2.2	Normalization of the Signals P, Q and S	32
4.3	Real Measurements	34
4.3.1	Additive White Gaussian Noise	34

4.3.2	Signal to Noise Ratio	34
4.3.3	Quantization Error	35
4.4	Measurement System	36
4.4.1	Existing Databases	36
4.4.2	Problem Statement	37
4.4.3	3-Phase Measurement System	38
4.4.3.1	Measurement Boxes	38
4.4.3.2	Switching and Detection Box	44
4.4.4	One-Phase Measurement System	45
4.4.5	Validation of the Measurements	45
4.5	Measurements	47
4.5.1	Appliances in Residential Buildings	47
4.5.2	Own Measurements	48
4.5.2.1	One-Phase Measurements Stand Alone	48
4.5.2.2	One-Phase Measurements Simulation	49
4.5.2.3	Measurements in Residential Buildings	50
4.6	Analysis of the Appliance Signals	50
4.6.1	Transient Response	50
4.6.2	P, Q and S of the Appliances	52
4.6.3	Steady State Behavior of Automates	52
4.7	Filtering of the Signals	55
4.8	Conclusion	57
5	Event Detection and Detector	59
5.1	Introduction	60
5.2	Problem Statement	61
5.3	Methods	64
5.4	Event Detection Algorithm	68
5.4.1	General Structure of the Event Detector	68
5.4.2	High-Pass Filter for the Event Detection	69
5.4.2.1	Approach of Hart	69
5.4.2.2	High-Pass Filters from Image Processing	70
5.4.2.3	The Gradient Operators	71
5.4.2.4	Laplace-Operator	73
5.4.2.5	Simulation of the High-Pass Filters	74
5.4.3	Complex Filter Structure	76
5.4.3.1	Approach of Canny	76
5.4.3.2	Improvement of Canny Filter by Perona	78
5.4.3.3	Simulation of Peronas Filter	81
5.4.4	Short Time Fourier Transformation	83
5.4.4.1	Simulation of the STFT	84
5.4.5	Classification Function	87
5.4.5.1	Threshold for the Classifier	88
5.4.5.2	Simulation of the Classifier	89
5.5	Results for the Event Detection	90
5.5.1	Experimental Procedure	90
5.5.2	Best Signal for the Event Detection	91

5.5.3	Results of the Event Detection Methods	91
5.6	Conclusion	94
6	Classification	97
6.1	General Description of Neural Networks	98
6.1.1	Biological Systems	98
6.1.2	Historical Overview	99
6.1.3	Applications for Artificial Neural Networks	100
6.2	Artificial Neural Networks	101
6.2.1	Perceptron	101
6.2.2	Transfer Function	102
6.2.3	Structure of Networks	102
6.2.4	Learning Algorithms	104
6.2.4.1	Learning Rule for the Perceptron	105
6.2.4.2	Learning Rule for the MLP	105
6.2.5	Design of a MLP	106
6.2.6	ADALINE	107
6.2.6.1	Structure of an ADALINE	107
6.2.6.2	Learning Rule ADALINE	108
6.3	ADALINE for the Classification	109
6.3.1	ADALINE for Parameter Estimation	109
6.3.2	Simulation	110
6.3.3	Results	112
6.4	MLP for the Classification	114
6.4.1	Features for the Classification	115
6.4.1.1	Train the MLP	116
6.4.2	Results of the Classification Method	116
6.5	Conclusion	118
7	Conclusion and Perspectives	121
A	Measurements and Measurement System	127
A.1	Signals for NALM	128
A.2	Measurements in the Laboratory	129
A.3	Measurements in Residential Buildings	131
A.4	Harmonics of the Current	132
A.5	Transient Responses of Appliances	134
	List of Figures	137
	List of Tables	143
B	Bibliography	145

1 Introduction

Contents

1.1	Motivation	2
1.2	Systems, Approaches and Applications	2
1.3	Research of the ReSP	4
1.4	Problem Statement in the Thesis	6
1.5	Structure of the Thesis	6

1.1 Motivation

Saving energy is one of the most important energy issues of our time. The worldwide energy demand has risen during the past years. As an example the consumption in the European Union (EU) has increased from 1999 to 2004 by 10.8% (Bertoldi and Atanasiu, 2006). In contradiction to this development the amount of available resources is decreasing. Furthermore, the world population is in the growth and accordingly the energy demand will increase during the coming years. This makes energy saving a necessity.

One way to achieve this is to influence and change the behavior of the human population (Bertoldi and Atanasiu, 2006). E.g. the private electricity consumption accounts for approximately 27% of the worldwide electricity consumption (International Energy Agency, 2008). At this point, energy can be saved. To adapt the users' behavior to the new challenges of energy saving it is necessary to provide them with a transparent report of their energy usage. From (Ehrhardt-Martinez et al., 2010) we know, that it is possible with a report to reduce the energy consumption by 9%-18%. With such a reduction of the energy consumption the carbon dioxide emission will also be reduced.

Currently the user of electric appliances in households only has access to information on the total energy consumption. A detailed report of the individual consumption of the appliances in the household is not yet available. The recently sold electricity meters, installed at the central entry point in the household, only provide a history of the total consumption. These meters cannot provide information on active appliances and their individual energy consumption (VDE, 2010). In order to save energy, users of the equipment need to change their consumption behavior. This requires more detailed electricity information. This system should not cost too much in order to increase consumer acceptance.

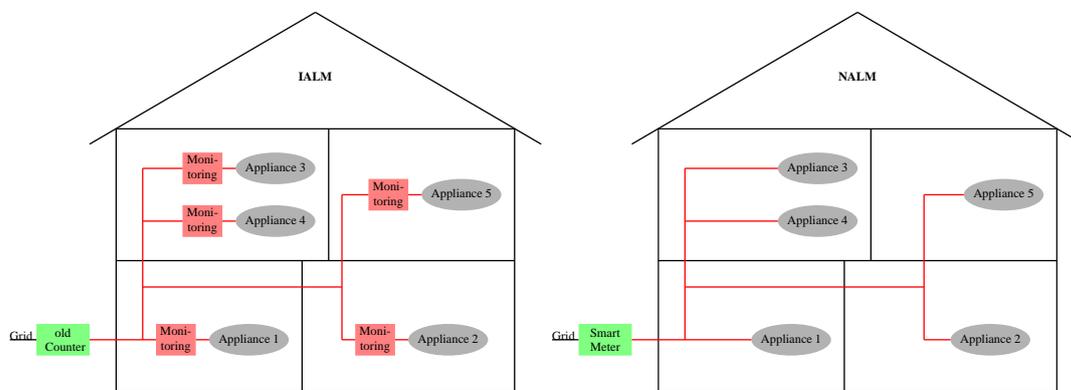
So it is important to develop a low cost energy measurement and analysis system. This system provides the consumer with more information. It is important that these information include how much and when each appliance consumes energy. This information could be presented in a detailed energy bill at the end of each month. The comparison with other appliances of the same type could reveal the out-dated equipment consuming too much energy. With that, the efficiency of the appliances in the household will be increased. To receive these information a monitoring system should be developed, which disaggregate the energy consumption of the individual appliances.

1.2 Systems, Approaches and Applications

There are different types of monitoring systems. Currently, many of the monitoring systems available on the market have to be directly connected to the appliance (digital-Strom, 2012). They measure the individual consumed power and transmit these data via a communication network such as the internet. This approach is called Intrusive Appliance Load Monitoring (IALM), because the electrical installation has to be modified for each appliance. The installation of the IALM technique in a household is

shown in Fig. 1.1(a). It is shown, that for each appliance a separate meter must be exist (digitalStrom, 2012).

However, IALM has many practical disadvantages. One disadvantage of IALM is, that the installation of the system requires considerable effort. To shorten the installation procedure and to reduce the costs another non-intrusive technique should be used. A system is required that is located at a central location in a household. The system must measure the entire (aggregated) electrical power consumption of the house. Therefore it is desirable to use only one measuring device at a central location point in each house, in this case at the building entry point. Using signal analysis algorithms, this signal can be separated into the individual load profiles of the appliances. This is called *disaggregation*. The methods and approaches in the field of disaggregation are described with the definition of Non-Intrusive Appliance Load Monitoring (NALM) (Najmeddine et al., 2008). The installation of the NALM technique in a household is shown in Fig. 1.1(b). This approach will be investigated in this thesis.



(a) Intrusive Appliance Load Monitoring (IALM). (b) Non-Intrusive Load Monitoring (NALM). For each appliance a separate monitoring system exist. These systems are plugged between the appliances and the sockets. That lead to a high installation effort. Only at a central point in the household the aggregate power consumption is measured. Disaggregation algorithms are disaggregating the energy consumption into the individual consumptions of each appliances.

Figure 1.1 – Two methods for measuring the energy consumption of electrical appliances in residential buildings.

NALM requires the measurement data from a central point in the home. Power meters measure at the entry point of residential buildings the total energy consumption. Previous systems, the old Ferraris-Meters, integrate the power consumption over the year and provide the customer with only one annual consumption bill. Through this integration, information is lost. In recent years, so-called *smart meters* have become increasingly popular. This now offer the opportunity to record the energy consumption in intervals of seconds. For the disaggregation the data of smart meters can be used. However, customers are currently reluctant to have such smart meters installed (VDE, 2010). This is mainly due to the costs involved. Furthermore, not many consumers realize the benefit of a smart meter. In order to give the consumers an incentive to get smart meters installed at their homes, it is necessary to clarify the advantages of the device. By the use of smart meters, further advantages arise.

Accurate data regarding the energy consumption patterns of appliances is essential if power supply companies want to manage power grid fluctuations more effectively. To obtain this data, smart meters must be installed in all households. Smart meters measure the data needed for the smart grid. The consumers' acceptance of smart meters is also essential for the success of smart grid. Smart meters are able to measure real and reactive power in one-second intervals. This information is crucial to electric supply companies. With this information the electric supply companies can compute the power factor ($\cos(\varphi)$) with higher precision than before. This way they can predict the load profiles of their appliances more accurately. With the prediction of the load profiles it will also possible to include the renewable energy resource more efficiently (Kazakidis et al., 2012). Another benefit is the faster and easier detection and localization of malfunctions in the power grid.

In addition, this system should help companies to optimize their load profiles and reduce energy costs. To adapt process control systems to cost-effective load profiles, for example, might be possible. By monitoring the energy consumption at a central location in the company, it will be possible to detect and predict malfunctions of large machines without stopping them. This will make it much easier to plan and adapt maintenance cycles to production processes. By dealing with large loads in industry, the term Non-Intrusive Load Monitoring (NILM) is often used. Since our approach is first applied to smaller household appliances, we will use NALM in the following text.

In general, there are two types of algorithms for NALM systems. The first one requires a database in which all possible appliances of households are listed. This is called *manual setup* (MS). The second system using *automatic setup* (AS). This system automatically gathers the information needed for functioning when it is installed. The research community is working on both types. There exists different approaches which use MS or AS. These are described in chapter 3. In this thesis a system with MS will be presented.

1.3 Research of the ReSP

The *Signal Processing Research Group* (RESP) is part of the faculty of *Electrical and Computer Engineering* of the *University of Furtwangen*. The research group has been working for four years with the solution of problems in the field NALM. In two research projects solutions for this area were developed. The issue of the research project *Smart Metering*¹ was the development and the implementation of the disaggregation of electrical appliances in residential buildings under contribution of NALM. In the context of this project, many works have been executed (Bier, 2010; Klein, 2011; Vogts, 2011). Another project is stronger address to the issues of analysis and visualization of the disaggregated data. This starts at the end of the year. In the field of algorithm development, the RESP is working with the research group of the TROP from *University of Haute-Alsace*.

¹Founded by the State of Baden-Wuerttemberg

The research area NALM at RESP can be divided into several fields of work. The main parts are the data acquisition, the event detection, the feature extraction, the pattern recognition, the tracking of the power consumption and the visualization of the results for the consumer. This chapter should give an overview of the whole research field at ReSP. Thus the reader should get a view of the wide field of smart metering.

The research has the goal to disaggregate the appliances in the load curve of residential buildings. For that, robustness algorithms should be developed. So the research field is divided in more parts (Benyoucef et al., 2010a,b, 2011a). Fig. 1.2 shows the structure of the research parts.

1. Creation of a database with measurements of individual appliances and measurements of households and companies.
 - ▶ Development of measurement systems for one phase/ multiple phases.
 - ▶ Verification of the electrical characteristic of the measurement system.
 - ▶ Store the measurements in the database.
2. To find the switching on- and off-events of the appliances an event detector should be developed.
3. Modeling of the appliances and feature extraction.
 - ▶ Using different time levels.
 - ▶ Creation of "finger prints" of the appliances for a first analyzes.
 - ▶ Analyze the behavior and the signals of different appliances.
4. Development of a classifier. This should classify the appliances which are switched on.
5. Development of a tracking algorithm which is used to calculate the power consumption of the appliances after they are classified.
6. The appliances in household can be divided into different groups. The merging of individual appliances to complex automates should be performed.
7. Results visualization. The results should be visualized that the consumer gets a transparent report of the energy usage.

At first, different measurement systems were developed to measure one phase and multiple phases. With those systems there were measured different appliances. These measurements were needed for the following development of the disaggregation algorithms. From measurements there will be extracted characteristics and models of classes from appliances. Based on the models of the appliances detection- and classification-algorithms will be developed. After the development the results are presented to the consumer. This thesis includes first topics of the research project which are described in the next section.

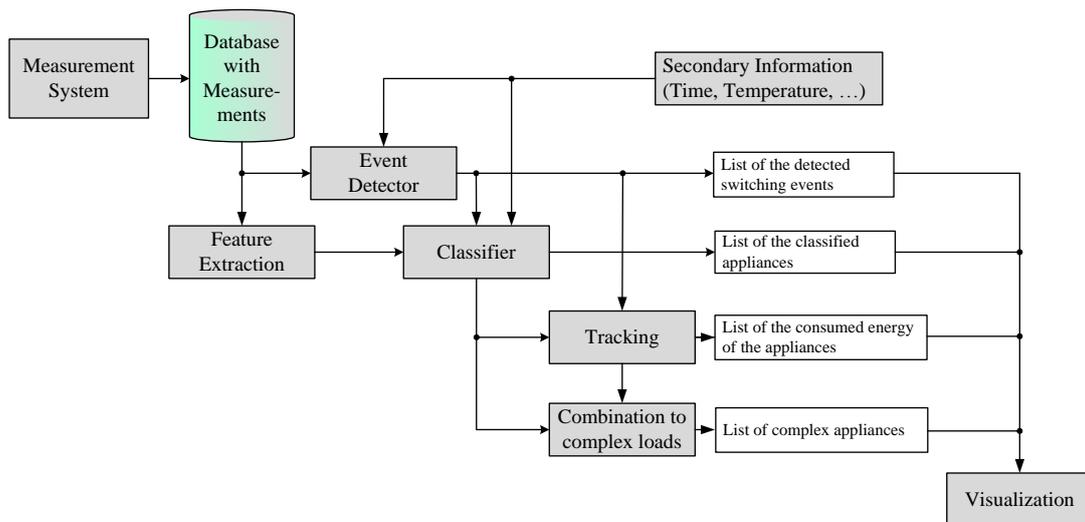


Figure 1.2 – Overview of the research area NALM at ReSP.

1.4 Problem Statement in the Thesis

In the previous section the research project *Smart Metering* was described. In this thesis, I worked on a part of the questions. The work in this thesis deals with the first issues.

1. Create Database with measurements of appliances in residential buildings. With these measurements the behavior of the appliances should be analyzed. Further these measurements should be used to develop and verify the developed disaggregation algorithms.
2. Development of an event detection algorithm to find the switching events of the appliances in the load profile.
3. Development of a classifier to classify a part of appliances.

The main part of that thesis is the development of an own measurement system, an event detection algorithm and a classification algorithm. The structure of that system is shown in Fig. 1.3. It is part of the whole disaggregation system illustrated in Fig. 1.2. The block *Measurement System* measures the entire signals of residential buildings. The signals are conditioned. It is possible to calculate other signals like active or reactive power from them. The recorded measurements are stored in a database. After that the block *Event Detector* detects the switching on- and off-events of the appliances. If a switching on-event is detected, the block *Classifier* is classifying the appliances. Both blocks, event detector and classifier create a list of their outputs.

1.5 Structure of the Thesis

This thesis is divided into seven chapters. After the introduction and the motivation of that work the research project is described. The approach of this thesis, which is

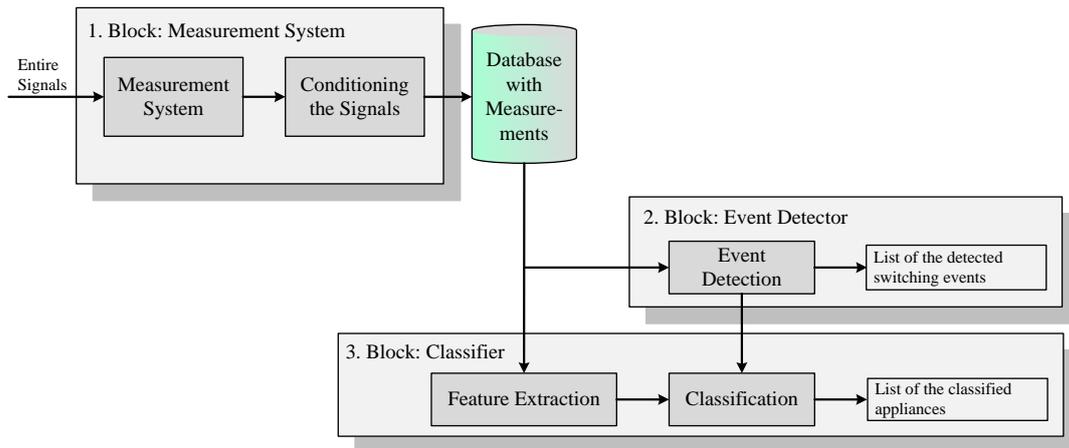


Figure 1.3 – Part of the system for the disaggregation of appliances. It shows the measurement system, the event detector and the classifier. Working on these three blocks is part of that thesis.

divided into three parts, will be shown.

Chapter 2 gives a short overview about signal processing and pattern recognition. The evaluation of classifiers is presented. For this several quality criteria are defined. Several types of classifiers are presented. It will be discussed which one is the best to solve the problem. At the end of the chapter, methods will be presented to analyze signals. It based on a time frequency analyzes. In the further chapter this method will be used.

The next chapter 3 describes the State of the Art of NALM Systems. The general framework for such a system is presented. It consists of the data acquisition, the feature extraction and the load identification. It will be shown an overview of some approaches of the last 20 years, where the first ideas for the NALM were founded. The usage of these approaches for this project is discussed.

Chapter 4 describes the own work of the thesis. It shows the signals, which are used for the disaggregation. A model of the appliances in the load profile is developed. Based on that model the signals are derived. In this work the power signals are used for the disaggregation. The measurement process with disturbances is described. After that the motivation to build an own measurement system is presented. After the system was developed and analyzed own measurements were carried out. These measurements are used for the algorithm development (Bier et al., 2013). The measurements were analyzed. Different features which are used in the following chapter are extracted. The chapter closes with the description of a filter structure. This filter is used to minimize disturbances from the grid.

The next chapter 5 describes the event detection as part of the disaggregation chain. The event detector is used to find the switching on- and off-events of the appliances. After the problem statement was defined, different methods were analyzed. Three approaches were developed. The first one based on a high-pass filter. The filter masks are derived from the gradient and Laplace operator. The second method using also a filter structure but with a more complex mask. The last algorithm uses a time

frequency method. All approaches are simulated and implemented. The detection rates for real measurement of the three algorithms were compared (Bier et al., 2013).

In chapter 6 a first method for the classification of appliance is presented. Different structures of artificial neural networks are used to classify the appliance. An approach is shown which use an ADALINE structure to extract the harmonics of the signals. This harmonics can be used to classify the appliances. A second approach uses different multilayer perceptrons. The ANNs are described. The simulation results are presented. At the end of the chapter the classification results using real measurement are presented (Bier et al., 2012).

The thesis is closed with chapter 7. It will be shown a conclusion and a perspective of the following work as part of the research project *Smart Metering*.

2 Pattern Recognition for NALM

Contents

2.1	General Definitions	10
2.2	Types of Classifiers	11
2.3	Evaluation of Classifiers	11
2.3.1	Confusion Matrix	12
2.3.2	Statistical Quality Criteria	12
2.3.2.1	Recall, True Positive Rate	13
2.3.2.2	Precision, Positive Predicted Value	13
2.3.2.3	Combined Measures	13
2.3.3	Determining a Rate	14
2.4	Time Frequency Analyze	14
2.4.1	Fourier Transformation	14
2.4.2	Short-Time Fourier Transformation	15
2.4.2.1	Transformation	15
2.4.2.2	Power Density Spectrum	16
2.4.3	S-Transformation	16
2.5	Conclusion	17

In this chapter an overview of the nomenclature in pattern recognition is presented. Pattern Recognition is the ability to find similarities, regularities, structures, repetitions and/or principles in a set of data.

2.1 General Definitions

The term *pattern recognition* was founded in engineering, whereas the term *machine learning* comes from computer science. Both terms describe the same. Both play a central role in more general areas such as *artificial intelligence* or *data mining*. Data mining extracts information from databases. It is the main part of regression, clustering and classification. *Regression* is a statistical method. The main goal of this method is to find relations between one depended and one or more independent variables. *Clustering* is an approach to recognize similarities in data sets. The founded groups are called clusters. If the individual groups are already known, the *classification* is used, to assign previously unassigned elements into a particular class.

We assume, we have measured a data-set \mathbf{x} . Now we like to develop a machine, which gets these vector \mathbf{x} as input. The machine should provide an output variable for different input vectors. Thus, better results are achieved in practical applications, an entire data set with N entries x_1, \dots, x_N is handed to the system. This is called the *training set*. It is used in most cases to adjust the parameters of the adaptive model. Suppose further that we have the target vector \mathbf{t} for each input vector. For example, this may have been determined from observations. The results of the learning algorithm can be represented as a function $y(x)$. The precision of this function is set in the training phase. This is also referred as learning phase. If the model is trained, new unknown data-sets can be identified. These are called *test-sets*. The ability, to identify new data-sets, which differs minimal from the training data, is called *generalization*. This is an important point in pattern recognition, because in practical use the training set can't represent all possible input vectors. One danger is the generalization adapting to the test data. This means, that the method used the test data memorize. This is called *over-fitting*. This needs to be counteracted in training phase, which is usually done with another data-set, used for *validation*. It is called validation data-set.

In practical applications, the data are preprocessed. These can be transformed into other spaces. This is done solving the problem there easier. The transformation is called *feature extraction*. Feature extraction is also used to increase the computational speed. In real-time applications individual features can be extracted from large data-sets. This reduces the complexity of the model and thus the computation time. So the goal is to find features that solve optimally the pattern recognition problem and provide a reduction of the computational effort.

In pattern recognition distinction is made between 2 learning methods. There are applications where the input vector and the corresponding output vector are known. These applications are called *supervised learning problems*. If the corresponding output vector are unknown this is called *unsupervised learning problems*.

In this thesis the disaggregation problem is solved using pattern recognition methods. To find the switching events of the appliances a classifier is used. This should classify

whether an event exist or not. The second classifier should classify different types of appliances. The learning phases are supervised.

2.2 Types of Classifiers

Classifiers consist of a decision function $f(x) = \mathbf{x}^T \mathbf{w}$ and a classification function $g(f(x)) = \text{sign}(f(x) + \Theta)$. In the most cases for the classification function a decision threshold maker is used. Linear problems can be solved using linear classifier. If the problem is non-linear, a non-linear classifier must be used to solve it.

As already mentioned, classifiers can be divided in types of their learning. Through the measurements performed in this work (cf. chapter 4) both input and output vector are known. For this reason only methods presented here, which are in the group of supervised learning classifiers. These methods can be divided into three groups.

The first group represents the parametric distribution based classifiers. They are based on the model assumptions regarding the characteristics of the pattern source, respectively on the probability density function. Is the function unknown it must be estimated. In the most practical cases a normal distribution is sufficient. Each distribution contains parameters which must be estimated from the sample pattern. A well-known representative of this group of classifiers is the Bayes classifier. This is based on the Bayes decision rule.

The second group represents non-parametric classifiers. These are model a Probability Density Function (PDF) without knowing the function in detail. The non-parametric classifiers include those who used for the classification directly the learning samples. A well-known classifier is the k-nearest neighbor.

The third classification category is based on function approximation. Parameters also have to be adapted for the training of function approximation classifiers. The key difference to parametric and non-parametric classifiers is that no distributions are modeled. Representatives are the polynomial classifier and the neural network classifier.

In this thesis classifiers should be used to classify appliances in the load profile of residential buildings. The PDF of the switching time and the working cycles of the appliances are unknown. For this reason a function approximation classifier should be used in this thesis. In chapter 6 this is described in detail. The approach uses an artificial neural network.

2.3 Evaluation of Classifiers

This section describes how the quality of a classifier will be statistically determined. The definitions here are based on the classification of events from appliances but can also be applied analogously to other systems of pattern recognition.

A classifier assigns patterns to classes based on certain characteristics. These patterns may be similar, but need not be exactly identical. If the patterns vary greatly from one another, the classifier will generally make errors. Thus assign a pattern to a wrong class

(e.g. for the event detection, detects an event where none is or analogous detected no event where one is). From the relative frequency of these errors quantitative measures to assess an event detector can be derived. There are two cases for detection: an event has been detected or not. The quality measures introduced here describe the binary case. In the following different quality criteria are presented.

2.3.1 Confusion Matrix

To evaluate a classifier, a number of cases must be applied. In which, at least, one should have in hindsight knowledge of the true class of the switching times. For the classification of an event four cases can be defined

1. True Positives (TP): An event exists and the classifier has detected it.
2. False Negatives (FN): An event exists but the classifier hasn't detected it.
3. False Positives (FP): No event exists but the classifier has detected one.
4. True Negatives (TN): No event exists and the classifier hasn't detected one.

In the first and fourth case, the classifier has detected properly appliances. In the other two cases, an error was detected. Now, for the evaluation of the algorithms the occurrence of the respective four cases are counted. For this purpose, a contingency table 2.1 (also called confusion matrix) is set.

	Event	NO Event
Detected	TP	FP
NOT Detected	FN	TN

Table 2.1 – Contingency table, for the example of the event detector.

Derived from the contingency table different quality criteria can define for the evaluation of the ED.

2.3.2 Statistical Quality Criteria

From the confusion matrix different criteria to evaluate the behavior of the classifier can be derived. For this purpose, relative frequencies are calculated. This contains the values TP , FP , TN and FN for each parameter value. To evaluate the ED, different evaluation measures are defined. These are defined for a specific threshold.

- ▶ Recall: $\frac{TP}{TP+FN}$
- ▶ Precision: $\frac{TP}{TP+FP}$
- ▶ Fallout: $\frac{FP}{TN+FP}$

For a first evaluation of the quality of our classifiers different rates are introduced (Klein et al., 2013; Anderson and Berges, 2012). These are specifically defined on the existing problem here. The individual decisions are now described in detail (for the example of the event detection).

2.3.2.1 Recall, True Positive Rate

We consider a detected event \hat{e}_j a *true positive* (TP) when there exists an e_i such that

$$\|e_i - \hat{e}_j\| < e_{thr}, \quad i = 1, \dots, N_e \quad (2.1)$$

with a time threshold e_{thr} . The threshold allows for correction of small delays between the measured data and the actually detected events. From the total number of detected events we get the number of true positives $N_{TP} \leq N_e$. The True Positive Rate (TPR) is the relation of number of true positives and true number of events

$$TPR = \frac{TP}{N_e}. \quad (2.2)$$

The TPR is therefore the recall applied to the TP. It answers the question: How many events are *right* detected? Another designation for the TPR is *Sensitivity*. This name is used for the evaluation of classifiers. The TPR is thus the conditional probability that an event has occurred and has been recognized by the classifier $P(e_j|e_j)$.

2.3.2.2 Precision, Positive Predicted Value

The PPV is the precision applied to the TP. It is calculated from the quotient of TP and the sum of TP and FP. The PPV gives a relation between right and wrong recognized event.

$$PPV = \frac{TP}{TP + FP} \quad (2.3)$$

2.3.2.3 Combined Measures

The precision and the recall only say in general not so much. For the detection of all events the TPR reached 100% and the detection of all relevant event the PPV reached 100%. Therefore, in some practical cases, the pre-defined two quality criteria are plotted against each other. For the determination a value of the decision function a specific threshold can be defined.

- ▶ Breack-Even-Point: threshold, where the Recall=Precision
- ▶ m F-Score: is calculated as the harmonic mean of precision and recall, $F - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$; $F - Score = 2 \cdot \frac{TP}{TP + FP + TP + FN}$.
- ▶ maximal F-Score: threshold where F-Score=max

2.3.3 Determining a Rate

The individual rates are mutually interdependent. This means that it is not possible to optimize all performance criteria independently. For this purpose, the two extreme cases of classifiers can be considered:

- ▶ *Liberal Classifier*, here almost all events are detected. The sensitivity is maximal. This, however, decreases the PPV, as many events are detected in the noise.
- ▶ *Conservative Classifier*, shows almost no events are detected. The parameters were chosen so that the PPV is very large. However, this has an impact on the TPR, which assumes lower values in this case.

The error FP , which makes the classifier, can be calculated from the TPR and PPV.

$$FP = \frac{N_e \cdot TPR \cdot (1 - PPV)}{PPV} \quad (2.4)$$

What type should be used by classifier thus depends on its application. The definition used here for the classifier and the associated choice of parameters are defined at the end of the chapter 5.5.1.

2.4 Time Frequency Analyze

One way to extract features from time signals is to transform the signal into the frequency domain. In this section a method is presented to transform the signals. This method is used later for the development of the event detection and classification algorithms.

2.4.1 Fourier Transformation

One known method to transform signals into frequency domain is the Fourier Transformation (FT). This is an integral transform and contains a harmonic oscillation $e^{-j2\pi ft}$ as integral kernel. It is a method of Fourier analysis. The FT allows it to transform a continuous, periodic signal $x(t)$ into its continuous spectrum. Thus, the similarity of the signal $x(t)$ with the harmonic wave of the frequency f is determined.

$$X(f) = \frac{1}{T} \int_0^T x(t) \cdot e^{-j2\pi ft} dt \quad (2.5)$$

This can be interpreted according to Parseval as a convolution of the spectrum with the Dirac impulse at the point $v = f$. In this theoretical assumption, the observation period is infinite. Therefore, in the spectrum no leakage effect occurs.

$$X(f) = \langle X(v), \delta(f - v) \rangle \quad (2.6)$$

Through the use of the entire signal, it is assumed that this is a time-invariant signal. For practical purposes in this work, this is a disadvantage. The Fourier integral cannot resolve non-stationary signals. Thus, changes in the waveform cannot be resolved. There must be found a method that dissolves non-stationary signals.

2.4.2 Short-Time Fourier Transformation

2.4.2.1 Transformation

Signals for the disaggregation are non-stationary in most cases. It is searched for a method which permits time-varying representation of the frequency distribution of a signal. Such a method is the Short-Time Fourier Transformation (STFT).

$$X(\tau, f) = \int x(t)w(t - \tau)e^{-j2\pi ft} dt \quad (2.7)$$

The continuous signal is multiplied by a window function w . The choice of a window function reduces the leakage effect. For the considered time interval the window shows values unequal 0. Known window functions are the Hamming-, Hann- or Gaussian-window. The values of the windowed time signal are then transformed analogously to the FT in frequency domain. The choice of window size affects the resolution in the time domain and frequency domain. However, the resolutions are disproportional. This relationship can be described about the uncertainty principle. That means, a resolution as high as possible in the time domain e.g. for event detection leads to a blurred resolution in the frequency domain. If one wishes to obtain a good frequency resolution, this again leads to a blurred resolution in the time domain. For the subsequent use of the STFT in this work, the choice of the window length has a great importance.

Fig. 2.1 shows the principle of the STFT. The result $X(\tau, f)$ of the STFT is a matrix. This contains the spectral values of the signal depending on the analysis time t and frequency f .

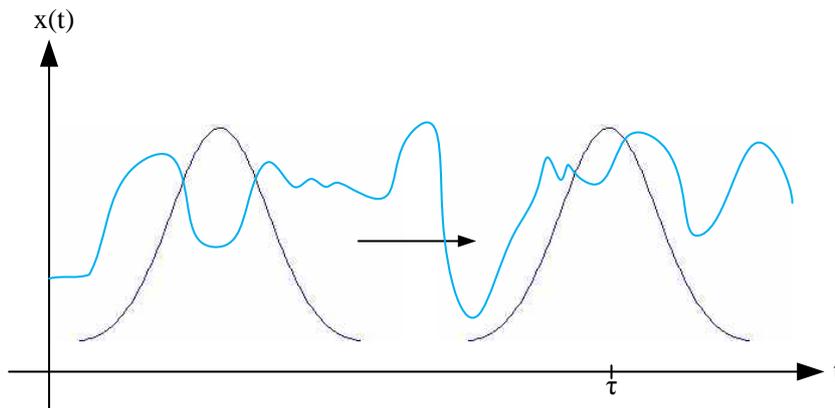


Figure 2.1 – Principle of the STFT. The signal is subjected stepwise at the times $t = \tau$ by a windowed FT. For each considered time, the FT provides all frequencies up to half the sampling frequency.

2.4.2.2 Power Density Spectrum

For a presentation and analysis of the results, the matrix X is limited. In most cases, the amplitude values are used for the analysis. For this, the individual components of the matrix must be squared. This represents the spectral density. When the spectral density is indicated over a specified frequency range, this resulting in the Power Density Spectrum (PDS)

$$PDS = \frac{1}{f_s} \cdot 2 \cdot \frac{1}{\omega_L} |X(\tau, f)|^2. \quad (2.8)$$

The three-dimensional representation of the PDS is called spectrogram. If we want to know the signal energy at a given time t , it must be integrated over the frequency components. Eq. (2.9) show this. The calculation of the double integral determines the total signal energy as presented in eq. (2.10).

$$\|x(t - \tau)\|^2 = \int PDS(\tau, f) df \quad (2.9)$$

$$\|x(t)\|^2 = \int \int PDS(\tau, f) d\tau df \quad (2.10)$$

In the following, an example is shown which should show the advantage of the STFT against FT. We assume there are two appliances, a linear and a non-linear. At the beginning the linear appliance is switched on. In the half of the measurement, at 0.5s the non-linear appliance is switched on. The non-linear appliance has to the fundamental wave (50 Hz) also some harmonics. These are 100 Hz, 150 Hz and 200 Hz. Fig. 2.2 shows the simulation results of the FT and the STFT. In the first row the time signal is shown. With the FT the whole signal is transformed into frequency domain. The results is presented in the second row. The fundamental wave and the harmonics of the second appliance are shown. But there is now information about the switching times. In the third row the results using the STFT is presented. The frequency components are plotted over the time. After 0.5s it is shown that a second appliance is switched on. Using the STFT it is also possible to detect the time of a switching event whereas the FT only calculates the harmonics.

2.4.3 S-Transformation

The S-transformation (also known as Stockwell-Transform) is a relatively new procedure in the field of time-frequency analyze. It was described by Stockwell first in 1996 (Stockwell et al., 1996). The S-Transformation (ST) is a modified form of the STFT. They differ in the window function. By the ST the window function is frequency dependent. The length of the window scales inversely with frequency.

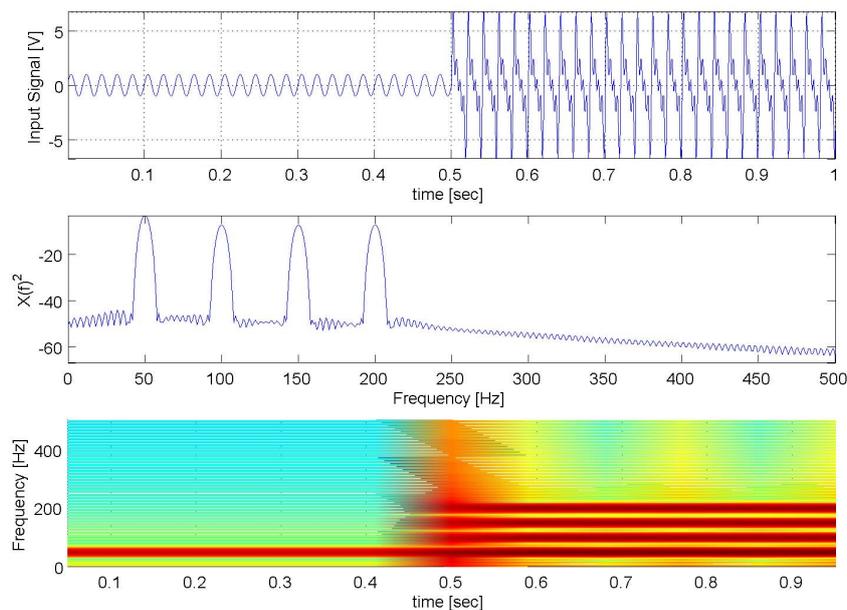


Figure 2.2 – Simulation results of the FT and the STFT for two different switched appliances. In the first row the input signal is shown. In the second row the FT of the input signal is shown and in the last row the results of the STFT is shown.

Using a frequency dependent window in the time domain, the ST is calculated for a time continuous signal $x(t)$ as

$$ST(\tau, f) = \int_{-\infty}^{\infty} x(t) \omega(\tau - t, f) e^{-j2\pi ft} dt \quad (2.11)$$

This equation corresponds to the known formula for the STFT (cf. (2.7)). But it has a window function, which is depending from the frequency f .

The ST looks promising as analysis method, but was not pursued in detail in the work. It will be shown later that the STFT for each method is sufficient. This method can be used for later work. Just when other signals are used as described in that work.

2.5 Conclusion

In this chapter the essential terms of pattern recognition are described. For the disaggregation of appliances a supervised learning method is used in this work. For this reason, an overview of the most common classifiers in this area is given. In order to make a statement about their quality, different criteria were defined. Based on the confusion matrix, the true positive rate and the positive prediction value were derived. Also combined measures are presented.

At the end of the chapter a method is presented with which features can be extracted. This is based on a time-frequency analyzes. Based on the FT the STFT is described. This method is used later in the work. A modification of the STFT has been presented for a further work. This is the ST.

3 State of Art for NALM

Contents

3.1	State of Research	20
3.1.1	Steady State	21
3.1.2	Transient State	23
3.1.3	Newer Approaches for NALM Systems	24
3.2	State of Technology	25
3.3	Conclusion	26

In the following chapter the state of art for Non-Intrusive Appliance Load Monitoring (NALM) systems will be described. It will be presented an overview of the most important disaggregation systems. A detailed description of the state of art for the different algorithms for the event detection and the classification will be presented separately in their chapters 5 and 6. Also an overview of the current available monitoring systems will be presented.

3.1 State of Research

As shown in chapter 1, the disaggregation of electrical appliances is divided into two parts, the Intrusive Appliance Load Monitoring IALM and the Non-Intrusive Appliance Load Monitoring NALM. For that project we use the NALM approach. An important criterion to distinguish between the algorithms, which are described in that chapter is, whether they need a learning phase or not.

The first approaches for a NALM system based on the investigation of the appliances in the steady state. The method is called *Steady State* analyze. There follows other approaches that examine the signals in the transient response. The methods based on the analyze of the transient response are called *Transient State* analyze. In both analyzing methods, the individual approaches have been divided for years (Najmeddine et al., 2008). At the beginning and during this work an intense research on methods for NALM was done. Due to the high research interest in this field, numerous research groups have deepened in this area worldwide in the last decade. Nowadays, a subdivision in the two aforementioned groups of the steady state and transient state is insufficient. There are methods that use a combination of both. Also, there are methods which use other methods of analysis. An overview of the methods for NALM is shown in Fig. 3.1. In the following, the individual methods are described in more detail.

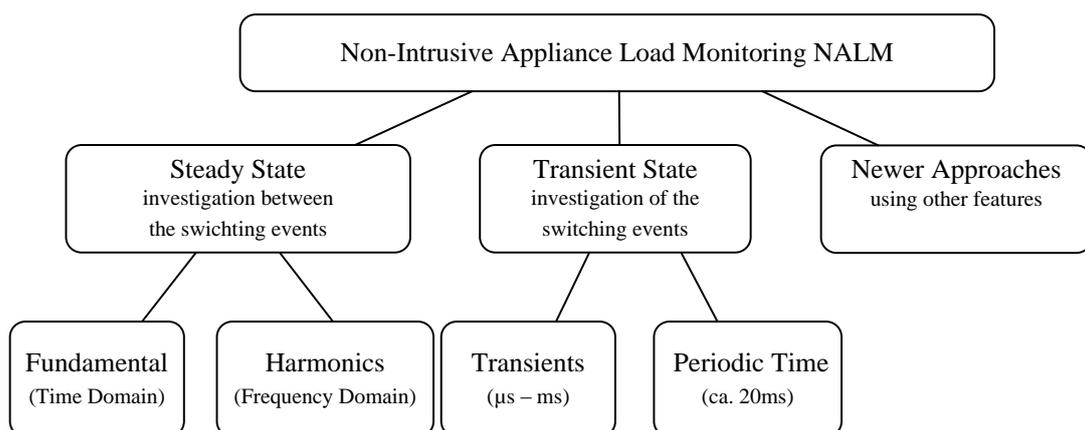


Figure 3.1 – Overview of Methods for NALM. The methods can be divided into three groups. The traditional ones are the Steady State and the Transient State analyze. In the last years other approaches are developed which deals not directly in these two fields. They can be placed in the third group.

3.1.1 Steady State

The first NALM-approach was developed by George Hart at the Massachusetts Institute of Technology (MIT). In the 1980s, he wanted to analyze the load of residential buildings (Hart, 1992). He had early discovered, that it must be possible, to detect the individual with the eyes visible switching events, with an algorithm. His system recorded active and reactive power in intervals of one second. After filtering, these results can be shown as a rectangular signal depending on the switching status of the appliances. The height of the steps in the power signal allows detection of the appliance. For that, before and after a switching event was measured. In that phase, the appliances were built up. Therefore the definition of the steady state was established. Thus the steady state analysis is before and after a switching event, which are described by transients. The analysis of the transients were neglected by Hart. For detecting multiple loads, it is important, that for each rising edge caused by switching on an appliance, a corresponding falling edge is recorded. An example of such a signal is given in Fig. 3.2.

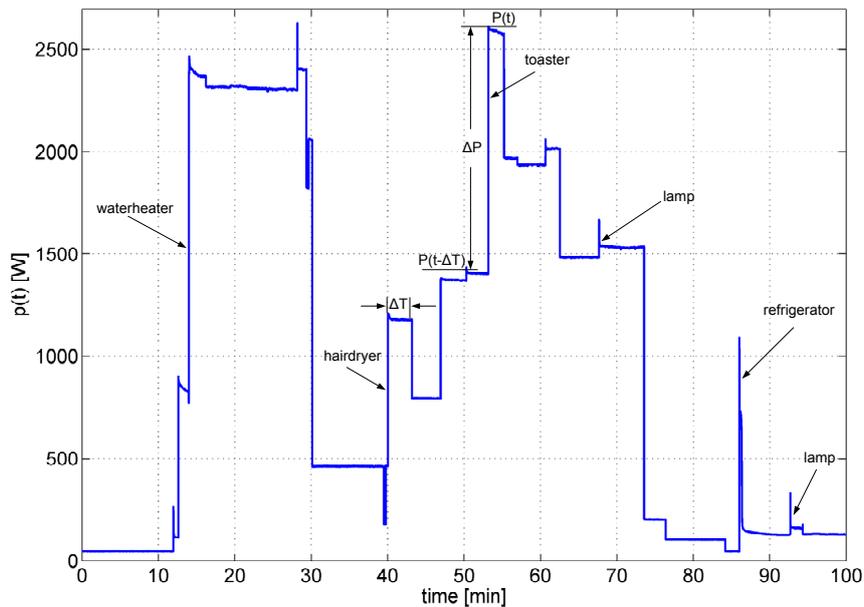


Figure 3.2 – Example of the power consumption of a household. Different appliances were switched on and off.

After the measurements of the appliances, the appliances were divided into different groups. Therefore, these were divided into clusters in the active and reactive space (PQ-Space). For each cluster of a switching on event, the point symmetric cluster of the switching off event were defined. In practice every switching event were allocated to the next center of a cluster.

The NALM approach of Hart has shown in field test, that only appliances with a huge power consumption are well classified. Furthermore synchronous switching events of linear and non linear loads are not classified. Also different appliances with the same load profile cannot be classified. A further disadvantage of this approach is, that

dependent from the low sampling frequency, fast switching events (phase angle control) cannot be detected.

The next approach here presented, was developed from Powers and Margossian (Powers et al., 1991) in 1991. They measured the load profile of residential building. With the aid of special information of appliances (e.g. periodical events of some appliances) they had generated characteristic load curves of appliances. For that the algorithm calculates the energy consumption of each appliance. Furthermore a-priori information of the appliances which located in the residential buildings are be collected. This process takes several years. The algorithms combine the measured values and the information of the appliances. The values were measured in intervals of 15 min. Thus that approach is vague related to the fast switching events.

1998 Pihala has developed a system in Finland, based on the approach of Hart (Pihala, 1998). It was a MS-NALM-System, which were implemented on a desktop PC. This system needed database values with the differences of the power profile of different appliances. In an interval of one second, 12 electrical variables were measured at each phase. In addition to the power differences, the start and end values of the power of each time interval were stored in the database. At the same time, the name of each appliance, which were connected on the phase were stored in the database. With that it was possible to classify more complex appliances (automate) with a predefined correctness.

The next algorithm needs at the beginning again some complex measurements. Marceau and Zmeureanu (Marceau and Zmeureanu, 2000) had analyzed seven large-scale appliances in Canada in 2000. Therefor they have measured for more weeks the apparent power of residential buildings on the two phases in Canada. After that, they have created a statistical profile of each appliance. The interval of the sampling rate was 16 s. The switching events were classified with the highest probability. The assumption was, that between that 16 s only one appliance was switched.

In Germany a system was developed in 2006. It based on the approach of Hart. It was developed by Baranski at the University of Paderborn (Baranski, 2006). The basis of his approach was the assumption, that the old ferraris counters are placed in the households for the next years. For this reason he built an optical measurement system. With that the velocity of the disc wheel were measured. With that method the instantaneous power can be calculated. Similar to the approach of Hart, that system has the disadvantage, that only a resolution in time interval of some seconds can be realized. Through the integration over the time of the signal, the transients of appliances cannot be detected. The group of appliances can be detected with that approach were similar with the group of the approach of Hart. With the approach of Baranski every detected event were clustered under the assumption, that every cluster has only one event. The center of each cluster was combined with the model of an appliance. Ideally should the sum of the switching events be zero, so the sequences were finished. For this reason the sum of the power differences of the switching on and off events of the appliances should also be zero. In addition a data vector for each appliance should be existing. For the identification of the models of the appliances he uses an generic algorithm. The analysis of the results of that approach has shown

some weaknesses. Baranski explains that result in order that it was not possible to distinguish between the appliances in the feature space.

To recognize simultaneously switching events of more appliances, it is possible to analyze the harmonics of the current. For that the time signal current were transformed with the Fourier Transformation in the frequency space. One work for that approach was founded 2005 in Japan. Nakano (Nakano, 2005) has developed a system to analyze the harmonics. The research group at the Central Institute of Electric Power Industry (CRIEPI) and the Tokyo University of Agricultural and Technology (TUAT) have measured the energy consumption of individual appliances and of residential buildings. With the created data set a support vector machine was trained, to recognize the events of the appliances. The power of the harmonics and the correspondent reactive values were used for the prediction. The results has an error of 20 % in comparison to the real measured values (Nakano, 2005).

Every of this six described systems needs a database, with information of the appliances. The installation and start up of those systems needs a high effort of measurements. All approaches were created for the classification of large-scale appliances.

3.1.2 Transient State

A second possibility to analyze the signals is the observation of the transients of the switching events of the appliances. Many disadvantages of the steady state analysis, like only the classification of large-scale appliances, can be eliminated by analyzing the transients of the switching events. When turning on or off an appliance, characteristic oscillations in the voltage and current signal may occur. The shape of those oscillations is dependent on the inner structure and the operation mode of the appliance. The distorted reactive power, the product of the distorted voltage and the distorted current, are mainly caused by non-linear loads, for example, switching power supplies. In contrast, linear but non-real loads like motors consume reactive power.

Lee established that the sum of the currents at higher frequencies can reach up to 150 % of the current in the fundamental wave of the power grid (Lee et al., 2003). Thus to determine the energy consumption of an appliance, it is crucial to regard the higher frequencies as well. Not only the current has harmonics. It is also possible, that the signal voltage has distortions. In fact, the distortions at the voltage are smaller as in the current but Cox et al. show in 2006, that the harmonics of the voltage are good to characterize the switching events (Cox et al., 2006). He uses the voltage between phase and zero potential and the voltage between zero-potential and ground. In the last case the higher harmonics are good for the classification.

Steven Leeb was the first one, who works at the classification using the transients of the signal (Leeb et al., 1995). Another approach using transients developed Shaw (Shaw, 2000). Building on this, in 2003 Lee (Lee et al., 2003) and Laughman (Laughman et al., 2003) developed a first approach. They use not only the steady state analysis. They use a combination of the steady state and the transient analysis to classify the appliances. For the simplification of that approach, it was assumed, that the signal voltage is an ideal sinusoidal. Thus the distortions of the voltage were neglected. Only the harmonics of the current were used. The approach consists of measuring

the signals with a sample rate of 8kHz and windowing the signal in time of one period of the fundamental wave. The overlap of the windows were 15%. For the US Grid with 60 Hz fundamental wave, 120 measured values per second were measured. Therefore the power of every harmonic get one value. With that approach appliances with similar active power consumption but with different harmonics were divided into different groups. For the classification of the transients of the appliances, a pattern recognition algorithm was developed. That algorithm based on the correlation of the switching on signal with all possible mean switching events of the appliances. With the measured values a regression with the mean values of the transients were calculated. The criterion is the gradient of the straight line. Is the gradient higher than one, then the two vectors are similar.

As mentioned before, Lee and Laughman had used a combination of the steady state and the transient state analysis. It was shown, that with this approach the classification was improved. This knowledge was also used by Lam et al. (Lam et al., 2004). Unlike as Lee, Lam has not developed a database with the different appliances. He has stored characteristic attributes of the load curve of different appliances in the database. For this a new taxonomy was developed. Which are better as the classical characteristic in that area (Lam et al., 2007). For that, the signals of the current and the voltage were compared for each period. Is the phase between these two signals similar, the bisecting line of the first quadrant comes out, for the resulting trajectory. If the signals have different phase angle, an ellipse will be created. Under the usage of semiconductors the trajectories gets additional distortions. Lam et al. has extracted characteristics from that signals, e.g. area, curvature etc. These characteristics were clustered. This clusters were better in reference to the characteristics of the appliances (Lam et al., 2007).

The last approach, which is here presented, was developed by Onoda (Onoda et al., 2000). He likes to recognize appliances with integrated electronic frequency converter. They have no similar power consumption. Thus they have no similar signals. Of total six appliances, the harmonics and the phase shift were measured. Different algorithms were tested. The best one was a boosting algorithm (Onoda et al., 2000).

The algorithms presented here based on the analyze of the transient or a combination of the steady state and the transient state. The approaches are developed for a small group of appliances. In the next section newer approaches are presented.

3.1.3 Newer Approaches for NALM Systems

2012 an approach for prediction of load curves was presented by Kazakidis et al. (Kazakidis et al., 2012). The prediction should be done at an interval of 1 h. As feature vector he used the active power, the temperature and the current time. For the prediction an artificial neural network was used. The system was tested by a small data-set but with good results. It is estimated only the total load consumption. A disaggregation which appliances used how much energy was not reached in this work. However, it shows that Artificial Neural Network (ANN) can be used for the disaggregation.

A work in the area of demand response was published by He (Dawei He et al., 2013) in 2013. He developed a hardware and software. For the disaggregation simple features are used. The feature space includes the RMS value of the current, the mean value of the power, the fundamental phase angle, the variation of the RMS and the 3rd and 5th harmonics. He uses a self organizing map (based on a artificial neural network) for the clustering of the appliances. But he didn't describe in detail the rates of the clustering. He analyzed only waveforms of individual appliances. Superimposed waveforms have not been studied. He admits to detect simple appliances. There are appliances with high power consumption specified. The individual appliances are well detected. The detection rates for individual appliances are given between 77 % to 99 %.

For the identification of load profiles in the smart Grid, Fernandes (Fernandes et al., 2013) used an ANN method last year. The approach is more in the field of system development. The group like to build a system, which send the consumer a monthly bill of the separated consumed energy of the appliances over the internet. For the disaggregation the group uses a simplified approach. The feature vector consists of the harmonics of the current up to the 15th harmonic. They tested the ANN classifier on measurement of 6 appliances with 252 individual measurements for the validation. The approach works well with a precision for each appliance up to 93 %. The approach would have studied for more appliances. This would achieve a generalization for the entire household.

Also, a recent approach represents the group of Paradiso (Paradiso et al., 2013), 2013. They also used an ANN for the classification and a feature vector with the power signals. It refers more to the communication interface as the disaggregation. Also, the classification rates are very low. However, this is also due to the choice of this method not fit the measured appliances. There are a variety of automate measured. For evaluating of the system a low data-set is used. This does not allow a good statistical analysis and a comparison of the approach.

An approach based on a unsupervised method is presented in (Pattem, 2012). The feature vector includes the power P . For the clustering a hidden-Markov-model is used. The method is not described in detail. For the verification of the approach the REDD data-set was used (cf chapter 4.4.1). So the approach was developed for low power signatures. The approach reach a disaggregation rate between 56 % and 67 % for the different households. This is very low for a reliable disaggregation system.

An approach for a real-time system is presented by Ruzelli 2010 (Ruzzelli et al., 2010). He used current transducers to measure only the current. With an idealized voltage he calculated as features the power and derived other values from this. For evaluation of the system 3 appliances were investigated. The integration interval is 1 s. For the device he gets a detection rate of 84 %. For the classification he uses an ANN.

3.2 State of Technology

Since the beginning 2010 it is statutory in Germany (Bundesministeriums der Justiz, 2005) to assemble Smart Meters in new buildings. Currently available systems on the market are transfer the values of the power in intervals of 15 min. Rarer are systems

that transfer the values of the power in intervals of seconds. Known smart meter producer are EasyMeter, Itron and Landis+Gyr. EasyMeter provides a Smart meter which can measures in an interval of 1 s.

The current systems are monitoring the whole energy consumption. The systems provides the power signals P , Q and S . For this reason, it is not possible to declare the individual consumption of each appliance. For a development of a system which can be included in current smart meter hardware the usage of the power signals should be preferred.

3.3 Conclusion

In that chapter an overview of the world-wide work in the research area NALM is presented. Since the beginning of the research in the 1980s, different approaches were developed. They can be divided classical into the steady state and transient state analyze. But in the last years many approaches were investigated which based on both or which uses other criteria for the disaggregation. Based on the methods own approaches should be developed. At the end of the chapter an overview of existing smart meter hardware is presented.

4 Analyse of the Signals for NALM

Contents

4.1	Introduction	28
4.2	System Model	28
4.2.1	Appliance Model	28
4.2.2	Normalization of the Signals P, Q and S	32
4.3	Real Measurements	34
4.3.1	Additive White Gaussian Noise	34
4.3.2	Signal to Noise Ratio	34
4.3.3	Quantization Error	35
4.4	Measurement System	36
4.4.1	Existing Databases	36
4.4.2	Problem Statement	37
4.4.3	3-Phase Measurement System	38
4.4.3.1	Measurement Boxes	38
4.4.3.2	Switching and Detection Box	44
4.4.4	One-Phase Measurement System	45
4.4.5	Validation of the Measurements	45
4.5	Measurements	47
4.5.1	Appliances in Residential Buildings	47
4.5.2	Own Measurements	48
4.5.2.1	One-Phase Measurements Stand Alone	48
4.5.2.2	One-Phase Measurements Simulation	49
4.5.2.3	Measurements in Residential Buildings	50
4.6	Analysis of the Appliance Signals	50
4.6.1	Transient Response	50
4.6.2	P, Q and S of the Appliances	52
4.6.3	Steady State Behavior of Automates	52
4.7	Filtering of the Signals	55
4.8	Conclusion	57

This chapter describes the measurement system and the measured signals which are used for NALM. At first a model is described. This model includes the appliances in the load profile of residential buildings which is a part of the grid. The measured signals are described. From them other signals can be derived. The data processing chain with the influences of noise is described. Thereafter, individual databases that exist today are presented. It is discussed how these can be used for this work. There follows a description of the self-developed measurement system and the analyze of them. For the later development of disaggregation algorithms own measurements were carried out. These are listed and analyzed. At the end of the chapter a method will be presented, which is used to prepare the data. It is a pre-filter of the signal using a non-linear filter structure.

4.1 Introduction

For the development of disaggregation algorithms measurements are required, whose statistic is known. In the last chapter it was described that a disaggregation can be divided into several blocks. As shown in Fig. 4.1 for the event detection an event detector is used. This is described in chapter 5. Chapter 6 describes the classification of the appliances in the load profile. However, before we can begin with the development of these two blocks, the signals for the disaggregation must be measured and analyzed (Benyoucef et al., 2012b).

For this reason, a measurement system was developed at the beginning of the work, which allows to record own measurements. These measurements are analyzed and used for the development of the event detector and classifier algorithms. If the measurements are known, then a statistical evaluation of the algorithms can be done. For this work, the signals are measured at the entry point. The physical signals which can be measured are the time continuous signals voltage $u(t)$ and current $i(t)$. From these two values other signals for the disaggregation can be derived.

4.2 System Model

At first, the model is described which present the appliances in the load profile of residential buildings. The derivation of various signals, which can be used for the disaggregation, is shown. Based on that model a normalization of the signals must be done.

4.2.1 Appliance Model

To describe the on and off cycles of the appliances, the model of a real voltage-source was taken. Fig. 4.2 shows such a source (extern grid) with different appliances (intern home).

The grid can be modeled as a voltage source, with a voltage $u_s(t)$ and a impedance Z_s . Z_s represents the source impedance and can occur physically by grid losses. The

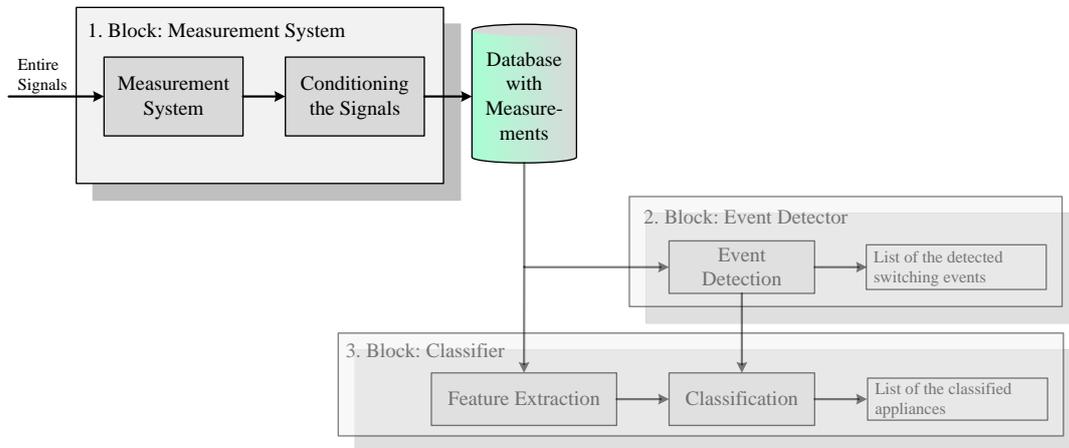


Figure 4.1 – Structure for the disaggregation of appliances in the load profile of residential buildings. The first block is described. The measurement system measures the entry signals $u(t)$ and $i(t)$. After the calculation of other signals and a pre-conditioning of them, these signals will be stored in a database. Later they can be used for the development of the algorithms for the event detector and the classifier.

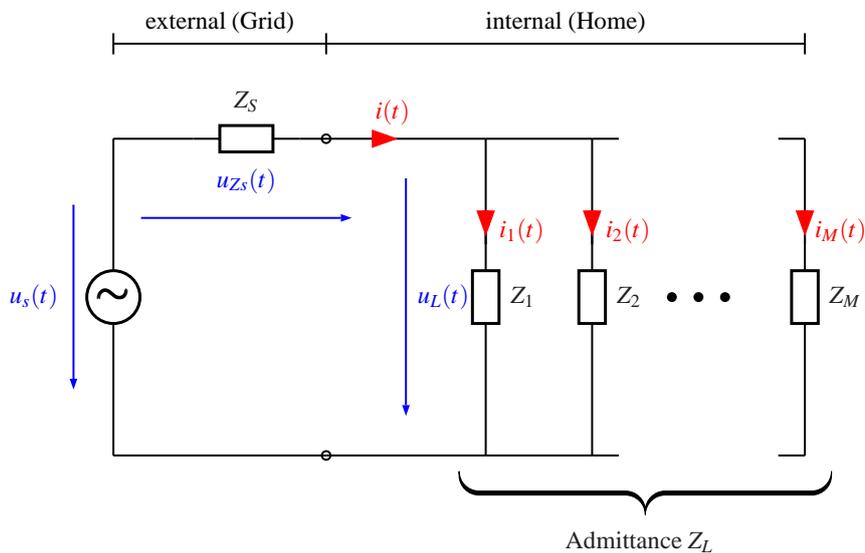


Figure 4.2 – Model for the appliances in the load profile of a residential building. The household is part of the external grid which can be modeled as a real voltage source and the resistance of the grid losses.

different appliances can be modeled as impedance Z_1 to Z_M where M is the number of active appliances. The total impedance Z_L of all appliances can be calculated as

$$Z_L = \frac{1}{\sum_{m=1}^M \frac{1}{Z_m}}.$$

The voltage across the appliances $u_L(t)$ can be calculated as

$$u_L(t) = u_s(t) - u_{Z_s}(t).$$

The total current of all appliances is calculated as

$$i(t) = \sum_{m=1}^M i_m(t)$$

For the development of the algorithms the following assumptions were made:

- ▶ It will be assumed, that the external voltage source is constant. There are no distortions in the grid and the effective value of the sinus with $f = 50$ Hz has 230 V.

$$u_s(t) = 230V \cdot \cos(\omega_1 t + \varphi_{u1})$$
- ▶ The second assumption is, that the external resistor of the grid is constant
 $Z_s = \text{const.}$

First, a linear resistive load is assumed. If a pure active resistance R or its inverse conductance G is connected to an AC voltage source with sinusoidal voltage $u(t) = \hat{u} \cos(\omega t + \varphi_u)$, it follows according to Ohm's law for the current $i(t) = \hat{i} \cos(\omega t + \varphi_i)$. These are the physical quantities which can be measured at the central point in the household. The measurement system shall capture and store these values.

Now, we know the basic signals which can be measured. Based on these, the signals we use in this work for the disaggregation should be derived. It has been reported in the state of art chapter 3 that many methods use the averaged values of the power signals. This has the advantage that a real-time system can be realized with a smaller sampling rate. As a result, it is possible to use existing hardware of smart meters¹ for a first check of the developed algorithms in a real case. At the beginning algorithms should be developed to get as input the power signals. The instantaneous power $p(t)$ of a linear resistive load is calculated from the product of the time-dependent current and voltage

$$p(t) = u(t) \cdot i(t) = \hat{u} \hat{i} \cos^2(\omega t + \varphi_u) = U I [1 + \cos(2\omega t + 2\varphi_u)]. \quad (4.1)$$

The power which is paid by the consumer is called active power. The active power P is calculated through the integration of a period from the instantaneous power². This leads to a down-sampling of the signal. Thus, the cutoff frequency of the signal is

¹e.g. EasyMeter model Q1D sample rate, 1 s

²Here is assumed that the signal is quasi stationary in a period.

change. The integrated signal P has a frequency of 50 Hz or a time period T between the measured samples of 20 ms.

$$P = \frac{1}{T} \int_0^T p(t) dt = UI = U_{eff} I_{eff} = I_{eff}^2 R = U_{eff}^2 \frac{1}{R} \quad (4.2)$$

Reactive power Q arises when the two-terminal circuit has inductive or capacitive elements. For a pure inductance, the time-dependent power results to

$$p_L(t) = \widehat{u}\widehat{i} \cos(\omega t + \varphi_u) \cos(\omega t + \varphi_u - \frac{\pi}{2}) = -UI \sin(2\omega t + 2\varphi_i). \quad (4.3)$$

Analogously results the power for a pure capacitance to

$$p_C(t) = UI \sin(2\omega t + 2\varphi_i). \quad (4.4)$$

So far, only the services of two-terminal circuit were examined which real power and reactive power added. In the following a system will be described, which can consist of a combination of R , C and L . This has the consequence that the phase difference $\varphi = \varphi_u - \varphi_i$ between the signals of current and voltage can take values in the range $-\pi/2 \leq \varphi_u - \varphi_i \leq \pi/2$. For the time-dependent power results

$$p(t) = UI \cos(\varphi_u - \varphi_i) + UI \cos(2\omega t + \varphi_u + \varphi_i). \quad (4.5)$$

By deforming the equation (4.5) one obtains

$$p(t) = UI \cos(\varphi_u - \varphi_i) [1 + \cos(2\omega t + 2\varphi_i)] - UI \sin(\varphi_u - \varphi_i) \sin(2\omega t + 2\varphi_i). \quad (4.6)$$

By integration the instantaneous power (4.6) over one period, we obtain for the active power

$$P = UI \cos(\varphi_u - \varphi_i) \quad (4.7)$$

and for the reactive power

$$Q = UI \sin(\varphi_u - \varphi_i). \quad (4.8)$$

Multiplies only the two rms values of current and voltage we obtain the so-called apparent power S . It corresponds to the hypotenuse of the power triangle, which can be calculated directly from the active and reactive power.

$$S = UI = \sqrt{P^2 + Q^2} \quad (4.9)$$

In the appendix A.1 is a summary of the most important signals for this work. The algorithms in chapter 5 and 6 gets these three signals for the detection and classification of the appliances.

Looking at non-linear loads, it is clear that distortions in the current are available. These distortions are called harmonics. By means of the Fourier Transformation can represent all periodic signals from the sum of individual harmonic functions. Thus, we are modeled the current $i(t)$ as a complex current consisting of the superposition of different harmonic currents. In this thesis, the harmonics are not used. Because we are used the integral signals P, Q and S as described later. Therefore, it is not discussed further here. During the development of our own measurement system the harmonics of the current will be analyzed. This is necessary because the measurement should also be used for further work. A description of the harmonics is in the appendix A.4.

4.2.2 Normalization of the Signals P, Q and S

Now we consider the model of the real voltage source closer. The resistor of the grid Z_S and the impedance Z_L of the appliance forms a voltage divider. This creates a dependence between Z_S and the power consumption of individual appliances. If more appliances in the household be actively pursued, the total parallel resistance Z_L will be lower. Thus, a higher current flows. The voltage drop across the parallel appliances decreases. Thus, the power consumption drops. From this, it follows that the real power consumption of the appliances varies depending on their number.

These voltage fluctuations can be considerable. Fig. 4.3 shows an example of the effective voltage for a real test pattern. As can be seen, the effective value of the voltage decreases if some appliances are switched on (e.g. at time 6 s, 11.5 s or 26 s). If appliances are switched off (e.g. at time 21.5 s, 31 s or 35.2 s), the effective value of the voltage is increasing.

This makes it clear that a normalization to the supply voltage (U_S) must be performed. Otherwise there is a risk that there will be unnecessary errors in the detection of the events or the subsequent classification of the appliances.

This problem was already described by Hart (Hart, 1992). The normalization should be represented formally for the example of the apparent power S . For the other two powers (P and Q) this is carried out analogously. For the supply voltage $U_S = 230V$ is assumed. This corresponds to the nominal value for the European grid. The normalized voltage calculated from the ratio of the squares of the supply voltage and the actual measured effective voltage U_L .

$$U_{norm} = \frac{U_S^2}{U_L^2} \quad (4.10)$$

With the normalized voltage a scaling of the power is carried out (a detailed description

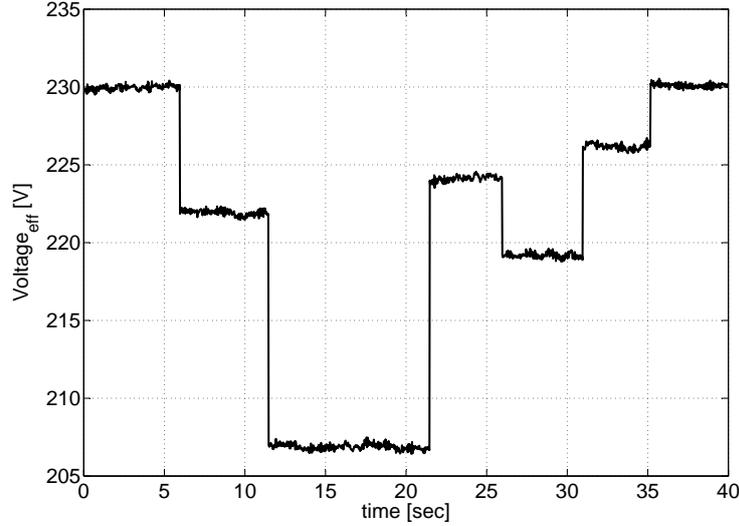


Figure 4.3 – Effective voltage across the appliances of a real measurement. Based on the voltage divider (Z_S and Z_L) the value of the voltage varies depending on the active appliances.

is found in (Hart, 1992)). This is calculated as

$$S_{norm} = S \cdot U_{norm} \quad (4.11)$$

$$S_{norm} = S \cdot \left(\frac{U_S}{U_L}\right)^2 \quad (4.12)$$

A computational example should clarify the effects of voltage fluctuations more in detail. Take the measurement from Fig. 4.3. After 11 s an appliance is switched on with a nominal power of 1985 V A. Together with the appliances already in progress the nominal power results in 2940 V A. With a constant voltage source, therefore, a current of 12.78 A would flow. As can be seen the effective voltage decreases approximately to 207 V. Accordingly, the actual power consumption reached a value of 2380.5 V A. It is missing about 630.39 V A from nominal power. This represents 19%. For a practical application, this value is too high. So it must be performed the normalization. With eq. (4.12) the power is normalized to

$$\begin{aligned} S_{norm} &= S \cdot U_{norm} \\ S_{norm} &= 2380.5 \text{ V A} \cdot \left(\frac{230 \text{ V}}{207 \text{ V}}\right)^2 \\ S_{norm} &= 2938.89 \text{ V A} \end{aligned}$$

The normalized power consumption consistent with the nominal value of the appliances now. The normalization was performed later for all measurements.

4.3 Real Measurements

Until now we have assumed that the measured signal is ideal. However, during a measurement errors will be occurring. These have different physical reasons.

The measured signals $u(t)$ and $i(t)$ are analogous time-continuous variables. These must be conditioned because the measured signal is changed by measuring. One possibility is that the measurements are additively superimposed by noise which comes from the measurement system. Another influencing factor is the processing of data in a PC. By the AD conversion, the continuous-time signal is converted into a discrete-time signal. The range of values is finite so quantization errors occur.

Our model for the real measurement is shown in Fig. 4.4. In the following, the two named factors will be further investigated. Their impact on the disaggregation is discussed.

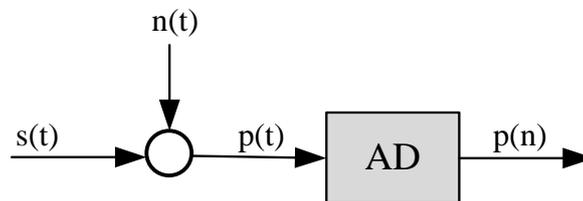


Figure 4.4 – Modell of a real measurement. The signal $s(t)$ are superimposed by noise $n(t)$. By digitizing a further change is made to the signal.

4.3.1 Additive White Gaussian Noise

Additive white Gaussian Noise (AWGN) describes a channel model. It has a constant noise power spectral density and is therefore called white. The signal amplitudes in the time level consists of a gaussian distribution. The signal $s(t)$ are superimposed additively by the noise $n(t)$. The measured signal $p(t)$ is

$$p(t) = s(t) + n(t). \quad (4.13)$$

White noise is produced e.g. by the random voltage across electrical components in a electrical circuit. Therefore, especially for the development of an own measurement system is to ensure a good noise cancellation.

4.3.2 Signal to Noise Ratio

In all technical applications useful signals are superimposed by noise. The intensity of the noise determines the technical quality of the signal. In the worst case it can lead that the useful signal doesn't provides a useful value.

A criteria for the evaluation of the superimposed signal with noise is the so-called signal-to-noise-ratio SNR. The signal-to-noise-ratio for an intensity signal can be defined as the relation of the signal power P_s and the noise power P_n over the same

system bandwidth.

$$SNR = \frac{P_s}{P_n} \quad (4.14)$$

The power loss over a resistance R is defined e.g. as $P = \frac{U_{eff}^2}{R}$. For the analysis of the system we assume normalized resistance ($R = 1\Omega$) so that we do not need to include that resistance term while measuring power or energy of a signal. We can then calculate the power of a periodical signal as the integral of the spectral power density of the signal $P_S(f)$ over the bandwidth B . If we assume white noise the power of noise can be calculated as the integral of the power spectral density $P_N(f)$ over the same bandwidth B .

$$SNR = \frac{\int_0^B P_s(f)df}{\int_0^B P_n(f)df} \quad (4.15)$$

In physics, signal amplitudes move frequently over several orders of magnitude. Therefore, it is sometimes useful to indicate the SNR in logarithmic form

$$SNR = 10 \cdot \log_{10} \left(\frac{P_S}{P_n} \right). \quad (4.16)$$

Eq. (4.16) is defined in general for a intensity signal. For other signal types, a further definition of the SNR must be chosen. In image processing the SNR of the image is defined as the quotient of the average gray value \bar{A} and the standard deviation σ of the gray value. It is also calculated logarithmically as

$$SNR = 10 \cdot \log_{10} \left(\frac{\bar{A}}{\sigma} \right). \quad (4.17)$$

For the disaggregation, the average gray scale value can be interpreted as half of the signal amplitude of the appliance.

4.3.3 Quantization Error

By an analog digital conversion, a time-continuous signal is converted into a time-discrete signal. Thereby the data processing in a computer is made possible. Because of the finite number of states the results must be rounded. This will cause error. These are referred as quantization error. The quantization error e_q is calculated from the real input value p and the quantized value of \hat{p}

$$e_q = p - \hat{p}. \quad (4.18)$$

The error cannot be larger than the quantization interval. In most cases, a linear quantization is selected with interval width Δ . If it is rounded to the nearest quantization,

the error is always between

$$-\frac{\Delta}{2} < e_q \leq \frac{\Delta}{2}. \quad (4.19)$$

The quantization error will occupy us later, when the entire measurement system is presented. The maximum quantization error indicates how small be the power consumption between the individual appliances may be to differentiate them.

4.4 Measurement System

For the development of a disaggregation solution and the later statistical verification of the system it is necessary to have multitude of measurements. In the literature a value of over 1000 events is listed (Bishop, 2006). In this section an overview of current available data sets is given. The usage in that work is discussed. After that the problem statement for the development of an own measurement system is presented. At last the development and verification of this is described.

4.4.1 Existing Databases

As already described, a system for the disaggregation should be developed, which should be implemented in a smart meter system later. For the development of the algorithms it can be possible to use data from a smart meter. Researchers however argue that most of the commercially available meters show a variation of 10% to 20% in data measurements (Zeifman and Roth, 2011). Additionally current smart meter have a limited hardware. Whereby a too low sampling rate is achieved. Therefore, for the development of a disaggregation system other data must be used. For this reason, several research groups have created their individual databases with an own measurement system.

In the previously described method (cf. chapter 3) it is usually not discussed in detail the data-set. But for a comparison of the different methods this is necessary. Some newer approaches from research groups already show data-sets with a good documentation. They provide them for a global development of disaggregation solutions.

Kolter and Johnsen presented a database called Reference Energy Disaggregation Data Set (REDD) in 2011 (Kolter and Johnson, 2011). They have measured different houses in USA. They differ based on the sampling rate between three types of data-sets. The raw data was sampled with $f_s = 15$ kHz. Therefor the voltage and current are measured. This high sampling rate allows a higher-frequency analysis of signals. But it also leads to a very high amount of data. For this reason only a small part of this is freely available. The second data-set is down-sampled from the raw data and for the voltage an ideal signal is assumed. The third data-set includes the power signals of 10 homes over 119 days. These data are sampled with $f_s = 1$ Hz.

In 2012, a further data-set was presented by Anderson et al. It is called BLUED (Anderson and Berges, 2012; Anderson et al., 2012). They have measured a household

with 24 individual circuits. So it is possible to compare the results with the individual power consumption of the appliances. It should be stated here that a lot of manual tuning effort was needed to generate the BLUED database. However, this database has some limitations because it has been shown that there exist variations between the time stamps in the documentation and the real switching events in the measurements. These variations make an analysis of the developed algorithms very difficult.

In 2013, a data-set was presented from Makonin for Canadian households (Makonin et al., 2013). The data were measured for one year and includes 11 measurements for 21 sub-meters. The measurements were measured with an interval of 1 min. This is very low for the development of disaggregation algorithms. Because of the low sampling rate, the transient will no longer show up correctly. This database can be used later if the power of appliances has to be calculated.

For a later work these data-sets can be used to compare the developed algorithms. Also the data can be used to compare the own measurements.

4.4.2 Problem Statement

At the beginning of this work measured data were not sufficiently available. Therefore, the development of an own measurement system was decided. This has also the advantage that the measurements can be carried out individually. By developing an own platform, it is possible to measure individual appliances. For this reason we describe a way of generating measurements with minimal manual tuning and data matching but with best possible match of the timestamps. The measurement system should achieve the following requirements

1. Development of a measurement system which can measure three and one-phase appliances. In residential buildings the most appliances have one phase. But some appliances like motors have three phases. It should be possible to measure both.
2. Measuring voltages with an effective value of 230 V.
3. Measuring of currents with a maximum effective value of 63 A. For the most residential buildings in Germany is this the nominal value at the entry point.
4. Measuring of appliances with a power consumption of 5 W. From chapter 3 we know approaches which deal with the disaggregation of appliances with a high power consumption. In this work, an approach should be presented which can also detect appliances with smaller power consumption. Small-scale appliances less than 5 W should not be disaggregated because their influence to save energy is negligible.
5. Optimum flexibility in measurement. It should be measured in control cabinets. Therefore, the wiring effort must be kept as small as possible.

In order to ensure greater flexibility, two systems have been developed. One is a single-phase measurement box. This was developed to quickly measure individual appliances. The second measurement box was designed for three-phase measurement. It is included

into a test bench. In the next section the test bench system with the hardware are presented. The measurement boxes are analyzed.

4.4.3 3-Phase Measurement System

The idea of the test bench system is to measure the energy consumption of appliances while recording their switching cycles. It consists of a measurement box (MB) used to measure voltage and current, a switching/detection box (SDB) used to control appliances and the storage of the measurement with a data acquisition card (DAC) and a PC. The block diagram is shown in Fig. 4.5. The components are described later. The system can be used in two scenarios:

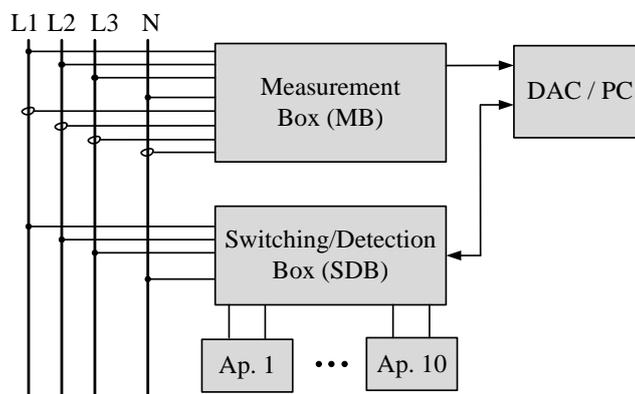


Figure 4.5 – Three phase test bench system. It is possible to simulate scenarios in the laboratory. It is also possible to record real measurements in residential buildings.

Simulation of Measurements With the system it is possible to simulate the switching events of appliances. Here the SDB toggles the appliances in predefined sequences. The measurement box continuously measures voltage and current. The measurements and the sequence are recorded to the PC. This scenario is intended for the laboratory because it allows for quickly generating a big data-set within a relatively short amount of time.

Real-World Measurements This scenario is intended in residential buildings. Here the appliances of the house have to be connected to the SDB. With internal sensors the SDB will detect whenever the owner uses a connected appliance and store this information on the PC. These real-world measurements depend on the behavior of the consumers in the household and are much more realistic than a simulated one.

4.4.3.1 Measurement Boxes

For a higher flexibility of the hardware, at the beginning of the work, two measurement boxes were developed (Benyoucef et al., 2010b,c). These boxes transform the grid

signals to measurable values for the DAC. Based on the act of requirement 1, that one- or three-phases must be measured, they should measure voltages and currents on one phase or up to three phases (L1, L2, L3) and the current of the neutral wire (N) of a residential building. From these signals it is possible to compute the secondary signals (e.g. P , Q , or S) for the disaggregation (Najmeddine et al., 2008). The safely measurable peak values are 325 V (against N) for the voltage and 89 A for the current³ defined in requirements 2 and 3.

The voltage is measured via a resistive voltage divider and a galvanic decoupling network to protect the measurement system from voltage peaks. The current is measured with current transducers. This has the advantage that the installation of the residential buildings does not have to be changed for the measurements. So requirement 5 is fulfilled. The block diagram of the measurement box for one phase is shown in Fig. 4.6.

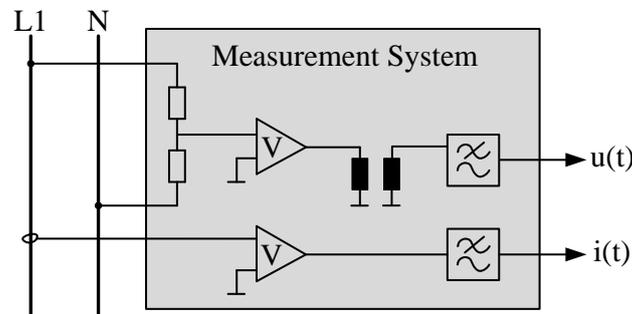


Figure 4.6 – Measurement principle of the measurement boxes for one phase.

To prevent aliasing both boxes have low-pass filters inside. For a higher flexibility of the measurements we have defined two different cut-off frequencies for the low-pass filters. The choice of box depends on the required sampling rate and the available storage space (memory). The parameters of the boxes are as follows:

- ▶ 3-Phase Measurement Box 1
 - ◆ voltage gain $G_v = 0.015 \text{ V/V}$
 - ◆ voltage cut-off frequency $f_c = 800 \text{ Hz}$
 - ◆ current gain $G_i = 0.26 \text{ V/A}$
 - ◆ current cut-off frequency $f_c = 1.3 \text{ kHz}$
- ▶ 3-Phase Measurement Box 2
 - ◆ voltage gain $G_v = 0.025 \text{ V/V}$
 - ◆ voltage cut-off frequency $f_c = 4 \text{ kHz}$
 - ◆ current gain $G_i = 0.15 \text{ V/A}$

³This corresponds to the effective values $U = 230 \text{ V}$ and $I = 63 \text{ A}$ which is suitable for most European households.

- ◆ current cut-off frequency $f_c = 4 \text{ kHz}$

We defined that the measurement system should detect appliances with an energy consumption greater than 5 W (requirement 4) which means that most appliances in the household can be measured. For this reason it is sufficient to detect current changes of 21.7 mA.

The signals are measured with the DAC⁴ with a resolution of 16 Bit. In the best case, this results in a resolution of $\frac{63 \text{ A}}{2^{16}} = 0.961 \text{ mA}$ corresponding to $0.961 \text{ mA} \cdot 230 \text{ V} = 0.22 \text{ W}$. So the maximum quantization error (cf. eq. (4.19)) is less than 0.11 W. Also appliances with an energy consumption below 0.22 W cannot be detected with this system. A second limit of the resolution is the current transducers. They have a lower limit in the area of some Milliamps.

Altogether the concept of the measurement boxes and the whole test bench meet the requirements 1-5. A photo of the two boxes are shown in Fig. 4.7. It follows the analyze of the transfer characteristic of the hardware.

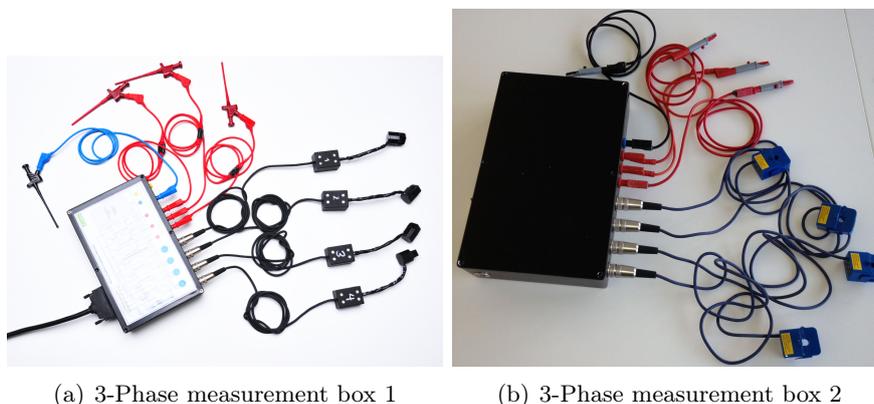


Figure 4.7 – Three phase measurement boxes. The signals $u(t)$ and $i(t)$ are conditioned for the AD-conversion. It is possible to measure the signals from all three phases.

The transfer characteristic of the measurement boxes consists of linear distortions modeled by the amplification G and the phase shift φ . The phase shift must be subtracted if the harmonics are analyzed. We assume that additional noise is superposing the signal. This noise comes from the thermal noise of the electronic components or the coupling of external disturbances. The noise of a system is modeled by the signal-noise-ratio SNR . Additionally there are nonlinear distortions which are parameterized by the distortion factor k . For the later verification of the algorithms it is necessary to know these characteristics of the boxes. With that knowledge it is possible to reduce the errors in the detection and classification algorithms.

For the analyze of the boxes the input and output signal should be know. The measurement procedure of the signals is shown in Fig. 4.8. For the evaluation of each parameter a defined signal $s(t)$ was applied to the input of the system. By measuring

⁴NI6361

the output signal $y(t)$, the parameter could be determined. As input signal, a sinusoidal signal is selected with different frequencies and amplitudes. The parameters are analyzed in the following.

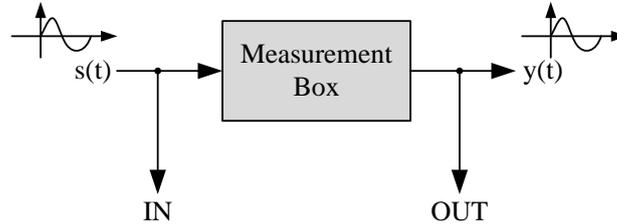


Figure 4.8 – Analyzing the measurement system. With the pre-defined input signal $s(t)$ and the measured output signal $y(t)$ the transfer characteristic of the measurement box can be determined.

Magnitude and Phase The first linear distortion is the frequency dependent gain. It influences the output of the box $Y(f)$ ⁵ by the multiplication of the input signal $P(f)$ (voltage or current) and the gain $G(f)$

$$Y(f) = G(f) \cdot P(f). \quad (4.20)$$

The second linear distortion is the phase shift φ of the input signal.

$$Y(f) = P(f) \cdot e^{j\varphi(f)} \quad (4.21)$$

The magnitudes of the phase were measured for the two boxes. The difference between the channels of the phases L1, L2 and L3 comes from the tolerance of the electronic components, which are used in the measuring box. As an example, Fig. 4.9 and 4.10 shows the Bode diagram of the first box for the voltage and current channels. The maximum of the voltage gain is approximately $G_{dB} = -36$ dB ($G = 0.01585$ V/V). The difference to the originally intended value of $G = 0.015$ V/V comes from the tolerances of the electronic components. The cut-off frequency is approximately at $f_c = 790$ Hz. For the current channels, the gain is $G_{dB} = -11.4$ dB ($G = 0.269$ A/V). The difference to the originally intended value of $G = 0.26$ A/V comes also from the tolerances of the electronic components. The cut-off frequency is approximately at $f_c = 1.3$ kHz (at a gain of -14.4 dB). The high-pass characteristic of the current channels comes from the behavior of the current transducers. It can be neglected because the cut-off frequency is lower as the frequency of the fundamental wave of 50 Hz.

As a second example Fig. 4.11 and 4.12 shows the Bode diagram of the second box's for the voltage and current channels. The differences to the calculated values are the same as for the first box. The voltage gain of the second box is approximately $G_{dB} = -32$ dB ($G = 0.0251$ V/V). The current gain of the second box is approximately $G_{dB} = -16.7$ dB ($G = 0.146$ A/V). Both channels have their cut-off frequency at 4 kHz.

⁵Signal in frequency domain.

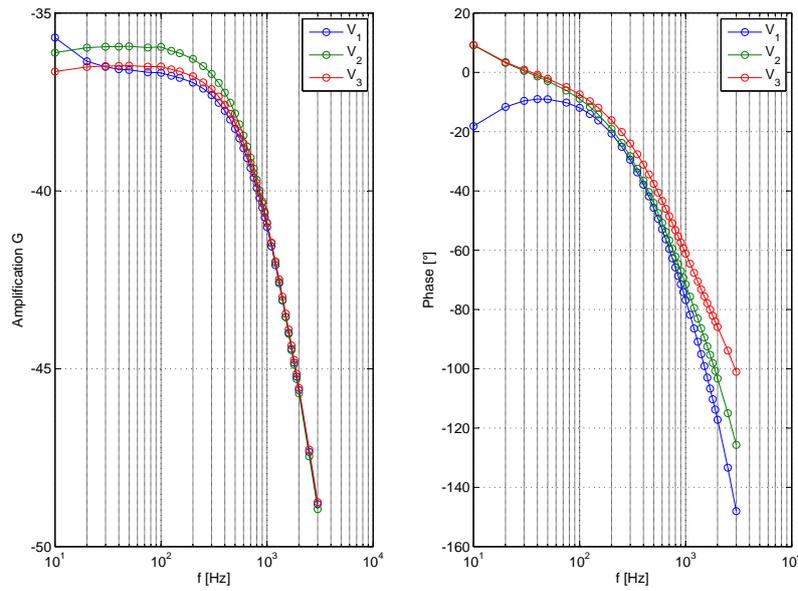


Figure 4.9 – Bode diagram of the voltage channels of the 1st measurement box

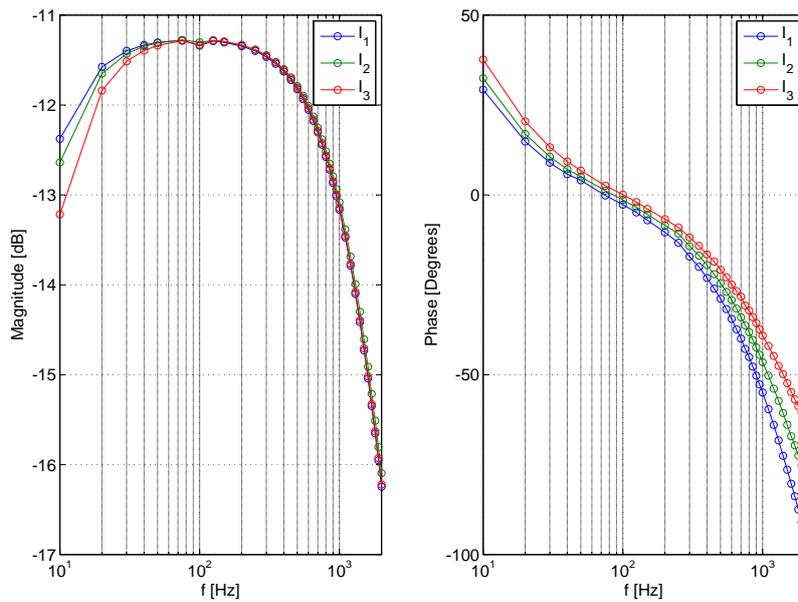


Figure 4.10 – Bode diagram of the current channels of the 1st measurement box.

Additionally for all channels the phase shift is presented. Because in this work only integrated signals are used for the disaggregation the phase shift can be neglected. But for further work using the fundamental signals it must be considered. So judging from the linear distortions both boxes are within their desired tolerance range of $\pm 10\%$. This value depends from the electronic components which are used.

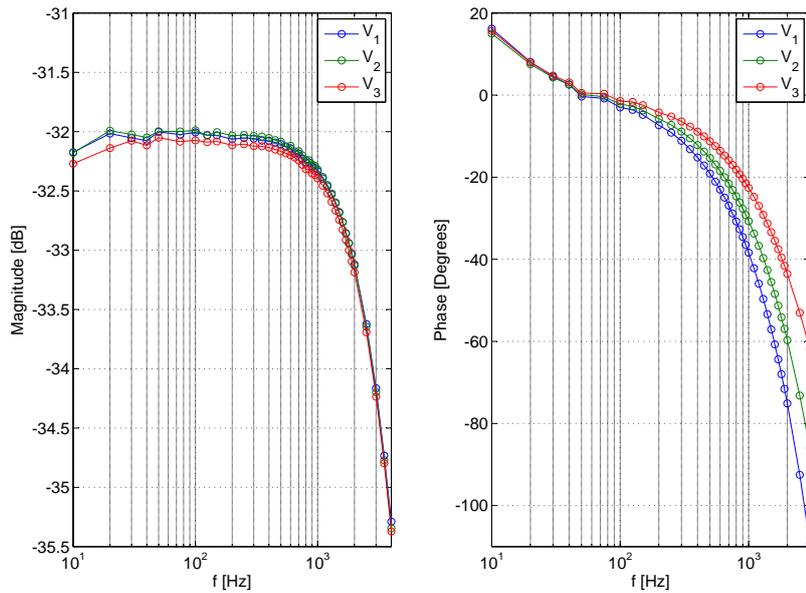


Figure 4.11 – Bode diagram of the voltage channels of the 2nd measurement box.

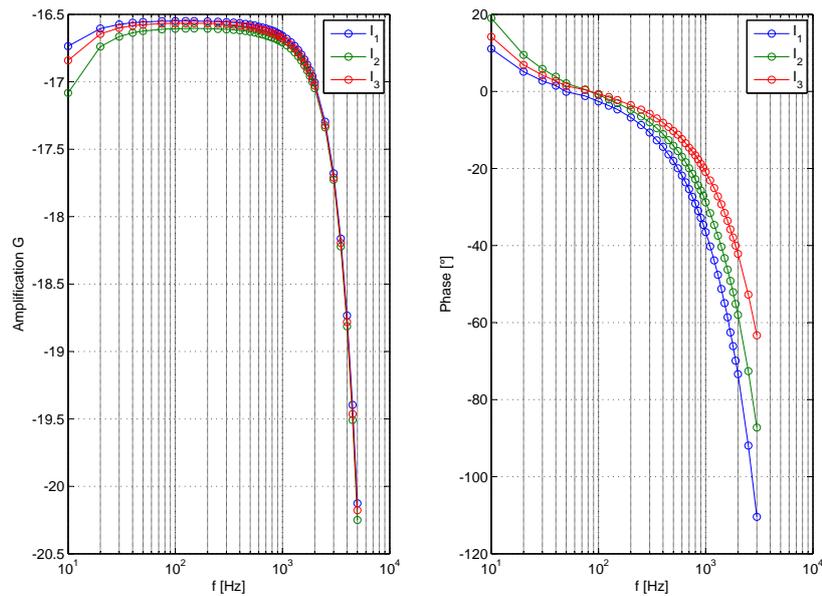


Figure 4.12 – Bode diagram of the current channels of the 2nd measurement box.

Nonlinear Distortion For a further work the fundamental signal $u(t)$ and $i(t)$ can be used for the disaggregation. Some approaches use the harmonics for the disaggregation. Nonlinearities in the electronics may cause distortions to those harmonics so it is important to know the behavior of the measurement boxes. A characterization of the harmonic distortion can be done with the distortion factor k

$$k = \sqrt{\frac{h_2^2 + h_3^2 + \dots + h_n^2}{h_1^2 + h_2^2 + h_3^2 + \dots + h_n^2}} \quad (4.22)$$

where h_x are the root mean square values of the harmonics. h_1 is the rms-value of the fundamental wave. The analysis of the boxes shows that the nonlinear distortion for the voltage channels is under 0.01 %. The distortion factor for the current channels is below 0.001 %. So they have no significant influence on the analysis of the higher harmonics of the input signals and can be neglected.

SNR The measurements have shown that the noise of the signal is white. For this reason the power of noise can be calculated as the integral of the power spectral density $P_N(f)$ over the same bandwidth B .

The SNR for the boxes are calculated with eq. (4.16). The bandwidth were approximately the frequency range between the 20Hz and the cut off frequency of the boxes. The SNR are shown in Table 4.1. It can be seen that the current channels have a better SNR than the voltage channels. This comes from the different electronic components. For the desired measurements the SNR values are sufficient.

Box	Channels	SNR
1	v(t)	> 15 dB
1	i(t)	> 31 dB
2	v(t)	> 15 dB
2	i(t)	> 33 dB

Table 4.1 – SNR of the two boxes

So all in all the analyze of the boxes have shown that these have no significant influence to the measurement. They can be used to measure the signals of the appliances and to create an own data-set for the development of the disaggregation algorithms.

4.4.3.2 Switching and Detection Box

The switching and detection box (SDB) is a fundamental part of the test bench system. The SDB has two main functions:

1. In the laboratory environment it can generate switching events for connected appliances. This means that it is possible to measure voltage and current while keeping track of the operational states of the appliances.
2. If the measurement system is used to evaluate the algorithms in the households, appliances are used by the inhabitants. In this case the SDB can detect the switching cycles with internal current sensors.

Two SDBs were developed. The first one is shown in Fig. 4.13(a). It SDB has 10 standard outlets with 16 A to connect up to 10 appliances. It is designed to draw an input current of up to 63 A. The second SDB is shown in Fig. 4.13(b). It has 8 standard outlets and one three-phase also with an input current of up to 63 A. To this SDB 9 appliances can be connected. To have a better flexibility by the measurements in the laboratory both boxes can be supplied with 63 A, 32 A or 16 A. For this different adapters were developed.



(a) Switching and detection box 1, consist of 10 standard outlets.
 (b) Switching and detection box 2, consist of 8 standard outlets and a three phase outlet.

Figure 4.13 – Switching and detection boxes.

4.4.4 One-Phase Measurement System

In the previous section, the test bench was presented. With that different scenarios can be measured. This test bench includes two three-phase boxes which can be connected directly to the entry point. However, since quick measurements should be carried out on individual appliances a further box was developed. This can be connected directly between an individual appliance and the power outlet. With this box, single-phase measurements can be performed on individual appliances. The electronics corresponds to the 3-phase measurement box 1. For this reason, in this thesis should not be discussed on the investigation of the behavior. Fig. 4.14 shows the one-phase measurement box. The conditioned signals analogous are converter with the DAC.

4.4.5 Validation of the Measurements

In section 4.4.3.1 the system parameters of the measurement boxes were analyzed. For the calculation of the individual parameters, the input signal must be known. For this, a sine wave with amplitude of 2 values was used. In the laboratory, only a sine wave with 1 A could be generated. This corresponds to a simulation of a consumer with about 230 W. However, in households, appliances with higher consumption are used (cf. tab. 4.3). For this reason the MS has been validated with reference measurements of a power meter. As a power meter, a device made by Fluke⁶ was used.

The Fluke measures the signals voltage and current directly and calculates the power signals P , Q and S and the harmonics of the current calculate. The calculated values

⁶Fluke 1760 Three-Phase Power Quality Analyzer



Figure 4.14 – One Phase measurement box. The box can be plugged directly between the appliance and the power outlet.

are stored in an internal memory. Depending on the sample rate, it can record data from 1 hour-1 month.

At the beginning of the measurement reference measurements were performed for the validation of the Measurement System (MS). Therfor in the Fluke was connected in parallel with the MS. Fig. 4.15 shows an example of a real measurement. The active power P is presented in the first row. Two appliances (with a great and a small energy consumption) were switched on and off. It is shown, that the Fluke and the MS measures similar values. The second row it was zoomed in. It is shown that the Fluke integrates over a greater period. The result is that fast changes are no longer displayed

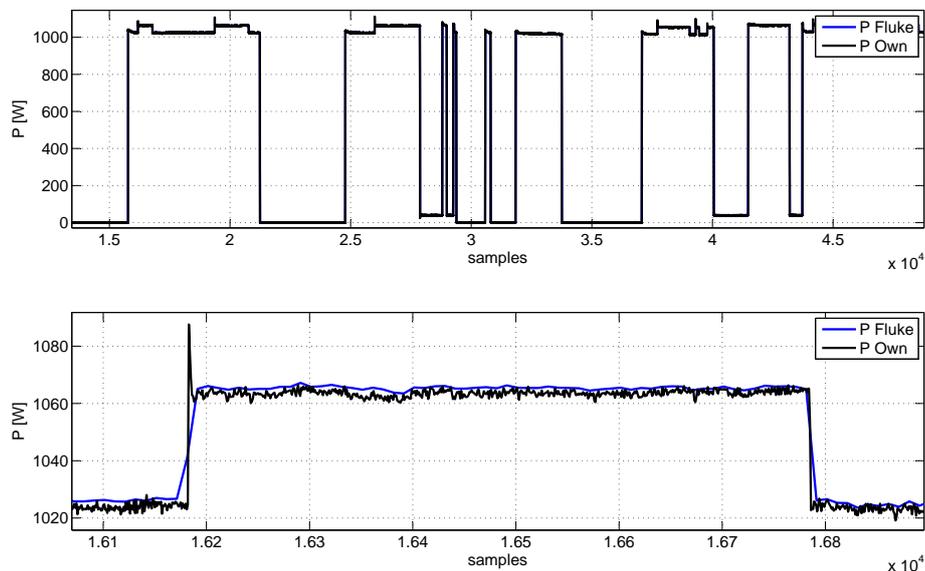


Figure 4.15 – Example of a measurement with the Fluke and the own developed measurement system.

Similarly, the power signals Q and S were evaluated. They have a similar behavior.

The validation of the MS with the Fluke power meter shows that the MS can measure also power consumption of greater appliances. After the validation of the MS, the measurements described below were performed.

4.5 Measurements

In the previous section the development of an own measurement system was described. With this system it is possible to perform own measurements on different appliances for the development of disaggregation algorithms. This database should be as representative as possible to cover the appliances in residential buildings. At the same time, a corresponding variety of similar appliances should be present. This is to investigate whether there are variations between the different appliances of a class. The database will serve as a reference database for the comparison of old and new algorithms. At the beginning separation of appliances into different classes has been performed. Depending on which class an appliance can be classified, can later draw conclusions about its switching cycles.

4.5.1 Appliances in Residential Buildings

Different measurements of individual appliances were done. The appliances can be divided into different groups of their behavior.

1. Classification of appliances according to their *switching states*.
Depending on the complexity of the appliances, these can be divided into groups of their switching states. Simple resistive loads (lamps, toaster, water heater ...) have two switching states. This can be easier detected, as automates. These groups can be divided even further. E.g. the automates can be divided into a group of autonomous setup (dishwasher, washing machine) and a group manual setup (TV, radio). Table 4.2 shows the appliances in the household, divided into groups of the switching behavior.
2. Classification of appliances according to their *power consumption*.
As described in chapter 3.1.1 many approaches describe the detection of major appliances. Accordingly, for future approaches it makes also sense to divide the appliances into groups of their power consumption. Tab. 4.3 shows the classification of the appliances into their power consumption in the residential building.
3. Classification of appliances according to their *local position*.
Appliances can be divided into groups of their local position in the house. By secondary information probabilities can be established whether an appliance is currently switched on or not. E.g. it is very likely if the light was turned on in the evening in the living room, then the TV is turned on.
4. Classification of appliances according to their *day use*.
A further classification of appliances can be made according to their daily use. In this case also a secondary information probabilities can be established. E.g.

it is more likely that the bathroom light in the morning is switched on, as in the basement.

Typ	Manual	Autonomous
1. on-off appliances, 2 states	lamps, toaster, water heater, PC-monitor	refrigerator, freezer, water pump
2. on-off appliances, 3 states	hairdryer, toaster with additional function	electrical oven (more states)
3. automates	laundry dryer, vacuum cleaner, mixer, oven, burner, washing machine, dishwasher, microwave oven, VCR, TV, coffee machine	boiler with several heating levels, air conditioner

Table 4.2 – Overview of the important appliances in residential buildings, divided into the type and the switching behavior.

Power Consumption (relatively costs)	Appliances
15.8 %	refrigerator, freezers, air conditioner
12.2 %	PC, communication electronic
11.5 %	water heater, boiler
11.1 %	light lamps
11.1 %	TV, radio, Hi-Fi
10.1 %	Drying
8.4 %	cooking
5.4 %	dishwasher
5.1 %	washing machine
9.3 %	others

Table 4.3 – Overview of the important appliances in residential buildings, divided into the type of their power consumption (NRW, 2006).

4.5.2 Own Measurements

Appliances should be disaggregated directly from the signal. The use of secondary information is not part of this work. For this reason the classification of appliances according to points 1 and 2 is investigated. For future work statistical information can be used to improve the monitoring system. For that the classification based on the point 3 and 4 can be used. Currently in the database there are measurements of individual appliances and pattern measurements.

4.5.2.1 One-Phase Measurements Stand Alone

In the presented databases (cf. chapter 4.4.1) are mostly only available measurements of total households. Individual measurements of individual appliances are only

few available. For this reason, the power consumption of individual appliances were recorded at the beginning of the measurements. Overall 400 measurements of individual appliances in 15 households were done. Same appliances differ for example in their size, manufacturer or date of manufacture. The appliances were measured with the measurement box described in chap 4.4.4. Based on the classification of Tab. 4.2 different scenarios were measured. For appliances of type 1 multiple switching cycles were switched manually. The same way was done with appliances of type 2. For complex appliances (automates) of the type 3 several functions have been run. E.g. for the washing machine different stages of the washing programme. A listing of all measured appliances is given in the appendix A.2. These measurement were used to analyze the behavior of the appliances and their variance between the individual classes. The analyze is presented in the following section 4.6.

4.5.2.2 One-Phase Measurements Simulation

A first set of measurements was collected for the verification of an event detection algorithm under laboratory conditions. The sequence files to switch the appliances on and off consist of a random sequence (uniformly distributed). The sequence were generated with switching times greater than 0.2s (this is the time delay between PC, DAC and the SDB). Since the measurements should be used for the development of the event detector, appliances from different classes (cf. Tab. 4.3) are used. The appliances were a water heater (with apparent power of $S = 953$ VA), a freezer (50 VA), a hand held mixer (75 VA), a hair-dryer (360 VA), a mixer (36 VA), an incandescent lamp (40 VA), a heater oven (1022 VA), an energy saving lamp (15 VA), a second hair dryer (599 VA) and a radio (3.8 VA).

Overall 10,000 switching events with the different appliances were measured. In chapter 5 and 6 these measurements are used to develop the event detection and classification algorithms.

The measurements were done in the laboratory. In the following some pictures of the measurement systems work are shown in fig 4.16.



(a) The whole system with different appliances. (b) Pc, ADC, measurement box 1 and switching detection box.

Figure 4.16 – Pictures of a performed measurement in the laboratory.

4.5.2.3 Measurements in Residential Buildings

During the period of the thesis measurements were performed in two households. The measurements were performed directly at the entry point of the households for all three phases. These are households with 5 respectively 4 persons. A total of 50 days were recorded. The measurements were performed for a first classification method (cf. chapter 6.3). This data is used to compare different methods for the disaggregation. Images of the measurement setup as well as some signal curves are listed in the appendix A.3.

4.6 Analysis of the Appliance Signals

For the later disaggregation it is to be examined, which features can be used for the classification. Based on the findings of chapter 3 and the analyze of own laboratory measurements different features were examined. With the analyzed data and the extracted features the disaggregation algorithms can be developed. In the following the features which have been investigated in this work are described.

- ▶ *Transient Response.* The switching on moment of the power Signals P, Q and S should be investigated whether these signals can be used as a feature to classify the appliances.
- ▶ *P, Q and S.* If appliances have the same instantaneous power it can be possible to use the active or reactive power to distinguish between these appliances.
- ▶ *Steady State Behavior.* In Tab. 4.2 different groups of appliances were defined. The complex automates should be analyzed whether it is possible to classify them using the power signals.
- ▶ *Harmonics.* The harmonics of the current signal can be used for the classification of the appliances. In this work only the integral signals P, Q and S are used for the disaggregation. So the harmonics of the current were not analyzed in detail. A short overview of the principle the harmonics is in the appendix A.4.

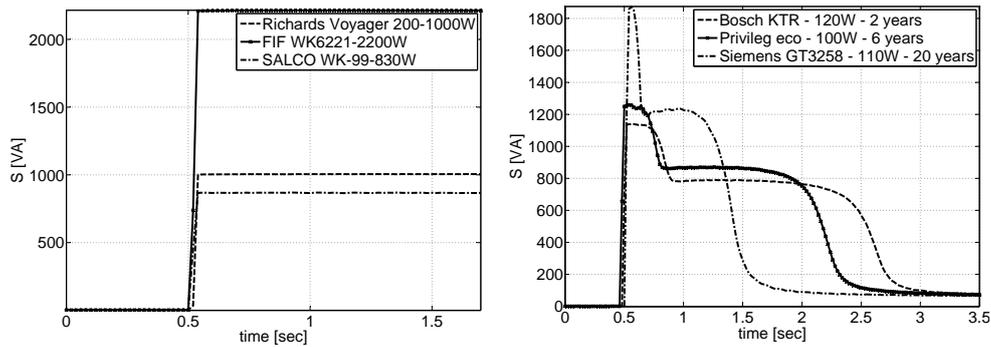
4.6.1 Transient Response

Each appliance takes some time until it has reached the steady state after switching on. In State of Art (cf. chapter 3) it was described, that the time of transient response is between 20 ms-3s. This time depends on the physical structure of the appliance. Easy on-off appliances of the first type are faster at steady state as complex automates. This is due to that the automates usually need to be initialized before operation. They also require more time to load the internal energy stores. Here is to be investigated how the transient response, based on the power signals, can be used for the classification.

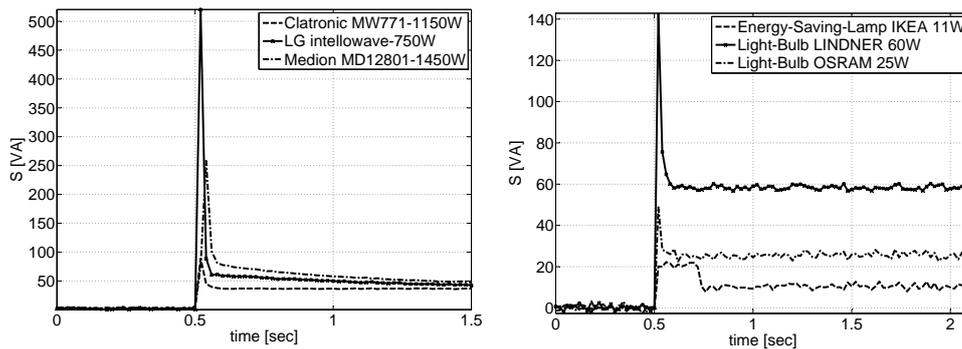
It should be investigated whether a generalization of the transient response of a particular type of appliances exists. For this purpose the measurements from chapter 4.5.2.1 were evaluated at the beginning of the work. In the following some examples

are discussed. Fig. 4.17 shows the transient response of 4 different types of appliances. After 0.5 s the appliances were switched on.

Fig. 4.17(a) shows the transient responses of several water heaters. These are among the first type the on-off appliances. The water heaters reach the steady state after 1-2 samples. There is also no overshoot. The shape of the transient response of all measured water heaters is identical. Only the power consumption varies.



(a) Transient response of several water heaters. (b) Transient response of several refrigerators.



(c) Transient response of several microwaves. (d) Transient response of several lamps.

Figure 4.17 – Different transient responses of appliances.

In fig. 4.17(b) the transient response of several refrigerators are shown. These have a distinctive transient. Which clearly differs from the water heater. At the beginning is an overshoot of approximately 200 ms. Then again comes a higher power consumption. This is between approximately 0.8 s-2 s. Then all devices reach their steady state. This behavior is typical for refrigeration and freezers. This is related to the internal structure of the compressor. The individual refrigerators differ only in their power. The shape is identically. It can be seen that the age of the appliances has no influence on the shape of the transient response. The 20 years old refrigerator has a similar shape as the 2 year old one. Thus, the behavior can be well generalized. The maximum amplitude is reached after 2 samples.

In Fig. 4.17(c) the transient responses of several microwaves is presented. The microwave is among the third type of appliances. The transient responses of the mi-

crowaves have a significant overshoot. But in comparison to the refrigerator this is very narrow. The maximum of the overshoot is reached after 2-3 samples.

The last type of appliances which is presented here are several lamps. The transient response is shown in Fig. 4.17(d). It will be distinguish between light bulbs and energy saving lamps. It is shown, that the light bulbs has a similar overshoot like the microwave. The energy saving lamp has a longer time a higher consumption. This comes from the internal power supply of the lamp.

The specification of the power consumption of the appliances is a nominal value. Here the transient responses of four different types of appliances are shown. So it is not possible to compare this value.

The transient response of other appliances is to be found in the appendix A.5. It has been shown that the transient response can be used to distinguish between the different types of appliances. It has been found that the transient response does not vary upon the age of the appliances or from the manufacturer. Only the amplitude values vary. The transient response based on the power signal can therefore be used as feature for the classification (cf. chapter 6).

4.6.2 P, Q and S of the Appliances

Another possibility to classify appliances is using the three different power signals as feature. Fig. 4.18 shows the active, reactive an apparent power of 4 appliances.

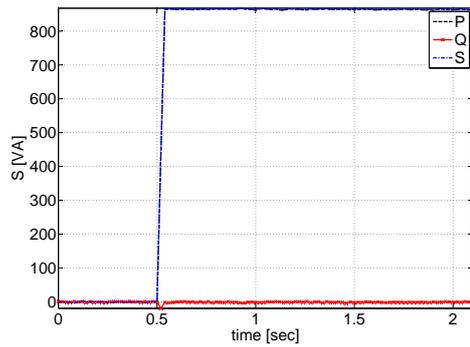
These appliance are in different groups (defined in Tab. 4.2). The water heater has only an active part of power as shown in Fig. 4.18(a). It is an ohmic consumer and has no reactive power. The power consumption of the light bulb is presented in Fig. 4.18(d). It shows that the light bulb has also only an active power. The microwave Fig. 4.18(c) and the refrigerator 4.18(b) has also an reactive power part. Q is positive so the appliances have an inductive behavior.

In Fig. 4.17 the transient responses of the same appliances were shown. It was shown, that the microwave and the lamp have a similar overshoot. Now we see that it is possible to distinguish between these two groups of appliances. If we have classify one of these to appliance we can use Q to verify the results. If it is near zero it is a lamp. But when it is unequal zero its more likely a microwave.

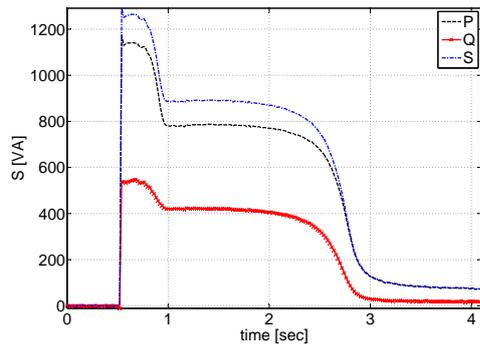
So the signal P, Q and S can be used for the classification. But for the event detector, developed in chapter 5 it could be a problem because the part of Q is for some appliance to small.

4.6.3 Steady State Behavior of Automates

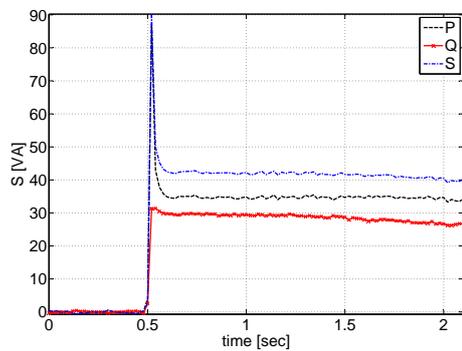
In Tab. 4.2 the appliances were divided into their switching states. There are on-off appliances, which may be switched manually or autonomously. In general, these have only two switching states. With the features previously described these types can be distinguish from each other. Fig. 4.19 shows the on-off cycles of a refrigerator and



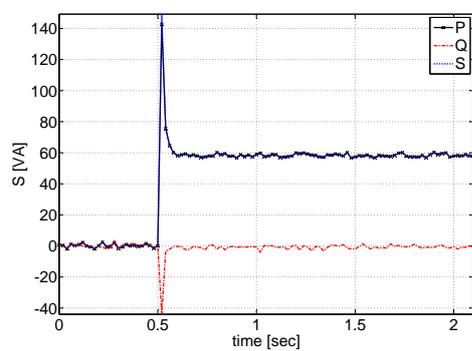
(a) P, Q and S of a water heater. P and S are dominant. Q has no influence.



(b) P, Q and S of a refrigerator. This appliance has all three power parts.



(c) P, Q and S of a microwave. This appliance has all three power parts.



(d) P, Q and S of a lamp. P and S are dominant. Q has no influence.

Figure 4.18 – Active (P), reactive (Q) and apparent power (S) of different appliances.

a lamp. The refrigerator was switched automatically whereas the lamp was switched manually.

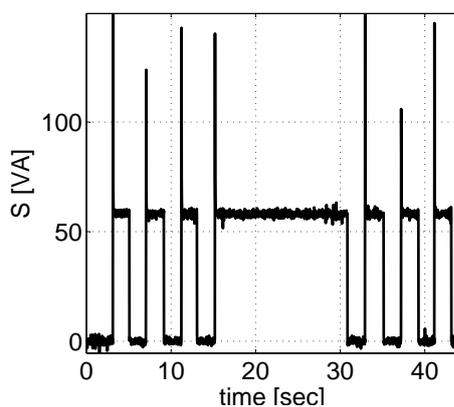
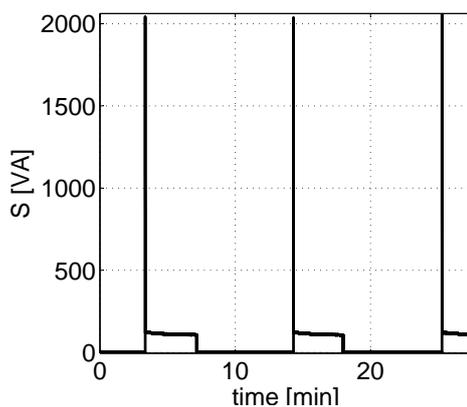


Figure 4.19 – Example of switching cycles from on-off appliances. On the left side the power consumption of an automatically switched refrigerator is shown. On the right side the power consumption of a manually switched lamp is shown.

Automates are not so easy to classify, because they have a complex internal structure. These consist internally again of simple on-off appliances. For example a washing machine consists of a motor, a heater element and the internal control electronic. Fig. 4.20 shows an operational procedure of a washing machine. Overall the machine works for 95 min. At the beginning after 5 min the heating phase is shown. The other cycles comes from the motor. Based on the previous described feature, the classifier would detect more different appliances.

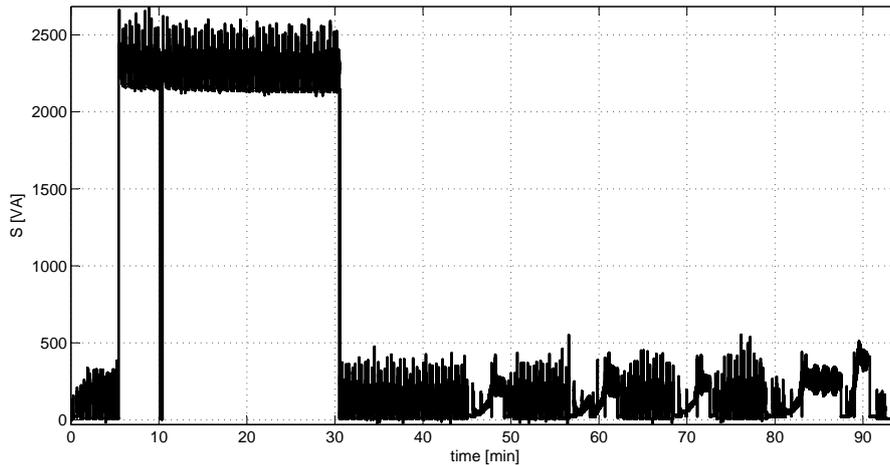


Figure 4.20 – Cycle-operating of a washing machine. It can be modeled as a combination of a water-heater, motor and control electronic.

Another example of an automate is shown in Fig. 4.21. It shows the power consumption of a dishwasher. The working cycle is 80 min. Several heating phases are shown. Similar to the washing machine, the classifier would detect different appliances.

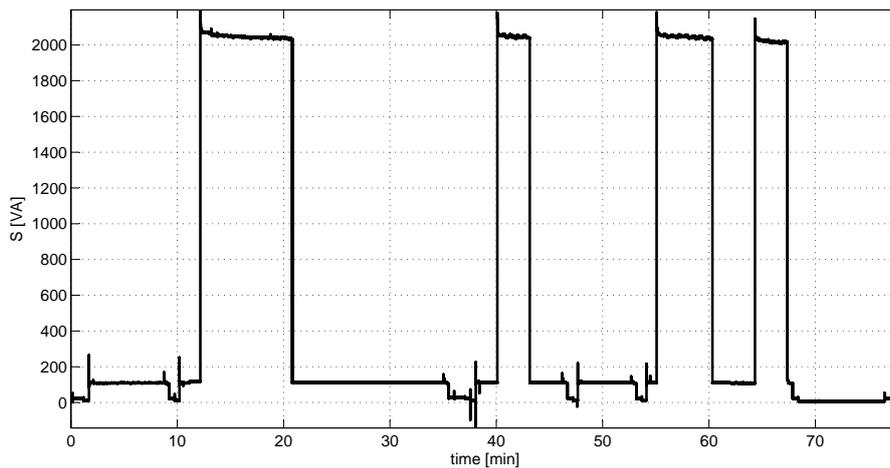


Figure 4.21 – Cycle-operating of a dishwasher. It can be modeled as a combination of a water-heater, motor and control electronic.

This work deals with the detection of events in the load profile. Also appliances should be classified in a broad load profile. The classification of automates will not be covered.

To classify these, models must be developed. For their development, the recorded and here presented measurements can be used.

4.7 Filtering of the Signals

From the analyze of the measurement system and pre-measurements we know that the measured signal is superimposed from different distortions (cf. chapter 5.2). At this point, a filter was used, which should pre-filters the signals. Especially the short distortions (peaks) from the grid should be reduced. This can later lead to undesirable errors in the disaggregation.

One way of noise reduction can be achieved by low-pass filtering. High-frequency components are cut above the cutoff frequency f_c . A low-pass filter is a linear filter. This leads also that frequency components of the original signal can be filtered out. This results in a smoothing of the switching on events of the appliances. Thus, the significant switching on events, which are used for the event detection and the classification, would be falsified. Therefore, a nonlinear filter was chosen for further work. This should be filtering the peaks from noise, but let the switching on event unchanged.

By averaging the signal it is possible to filter them. In statistics, different mean values for the distribution of a signal exist. These are

- ▶ *Arithmetic Mean*, is a mean value calculated from the quotient of the sum of all measured values and the number of them.
- ▶ *Mode*, is by an empirical frequency distribution the most common value,
- ▶ *Median*, is a mean for distributions in statistics. It is the value which stays in the middle of the measured values when they are sorted by their size.

The arithmetic mean is linear and for the pre-defined requirements does not used. The mode is the expression with the highest probability for a discrete random variable. Since only a few values are available for filtering (milliseconds between each event), the mode is not suitable here. The median is not linear. It is robust against outliers in the signal and is among others in the image processing as median filter to sort gray values in signals. The median filter is to be used here to filter the signal.

The filtered signal $P'(n)$ is calculated from the power signal $P(n)$ to

$$P'(n) = \text{median} \{P(n, n + 1, \dots, n + (m - 1))\} = \tilde{n} \quad (4.23)$$

where m is the number of values over which the signal is filtered. For different m , \tilde{n} becomes

$$\tilde{n} = \begin{cases} \frac{n_{m/2} + n_{m/2+1}}{2}, & \text{if } m \text{ even} \\ n_{m+1/2}, & \text{if } m \text{ odd} \end{cases} \quad (4.24)$$

By an increase of m , the signal will be filtered more. Fig. 4.22 shows a simulation of the median filter with different values of m . In the first row the input signal is shown. It is a peak at sample 3. In the second row the output of the median filter is shown. In the first column the median filter with $m = 1$ is shown. There is no change in the

signal. In the second column the median filter with $m = 2$ is shown. The value of the peaks is no halved but over two samples. In the last two column the outputs of the median for $m = 3$ and $m = 4$ are shown. The peak is completely filtered now.

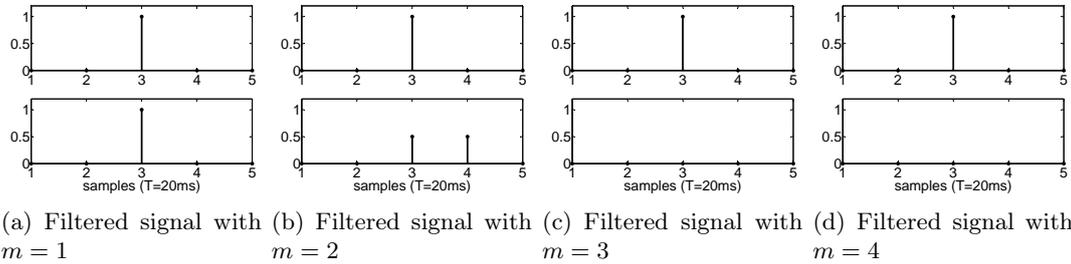


Figure 4.22 – Simulation of a median filter. The first row presents the input signal. In the second row the median filter outputs are shown.

What is the influence of the median filter for real changes in the signal like a switching on event? To answer this question a second simulation was done. In Fig. 4.23 the behavior of the median filter with different m of a step was analyzed. It is shown that the median filter with an odd value for m has no influence to the step. The median filters with an even value like $m = 2$ or $m = 4$ are change the input signal. The first value of the step is halved. From the first simulation we know, that the value of m must be greater than $m = 2$ to reduce faults from peaks with one sample. If a median filter is used with an even value of m the gradient of the switching on event is changed. From these reasons the optimal value of m based on the simulation is $m = 3$;

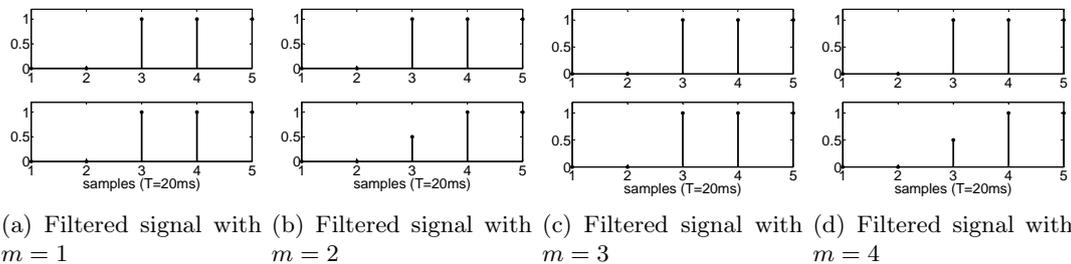


Figure 4.23 – Simulation of a median filter. The first row presents the input signal. In the second row the median filter outputs are shown.

The median filter was implemented and the measured signals from 4.5 were filtered. An example is shown in Figure 4.24. Here, a power signal from a real measurement was filtered with median filter with different values $m = 1,2,3,4$. It is shown, that the peak from noise at 105 s is decreasing but the events from the appliances (e.g. at 190 s and 225 s) are not filtered.

The median value at point n is calculated from the values of n until $n + m - 1$. This leads to a shift of the original signal by $m/2$. That means, the detected events are shifted of $m/2$ from the point n . This offset (which in the median calculation used here is minimal, because only small m are used) must be considered in the subsequent calculation.

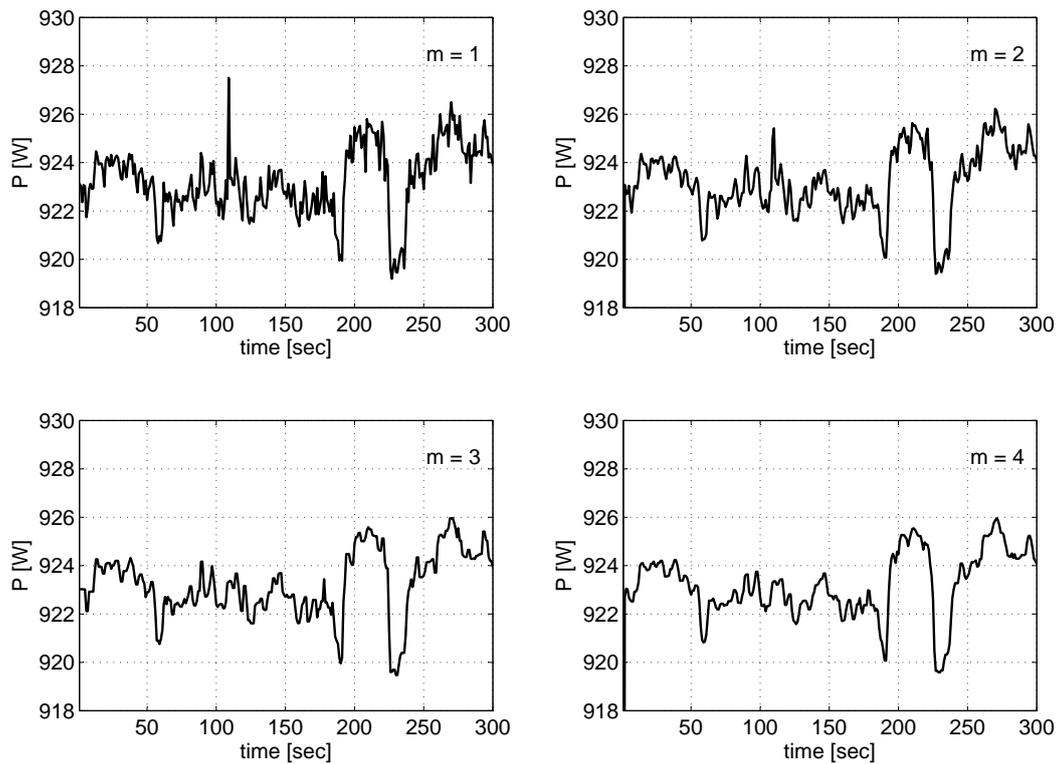


Figure 4.24 – Filter characteristic of the median filter. Example of a real measurement with a peak as distortion and different switching events of appliances.

4.8 Conclusion

This chapter discusses the first part of disaggregation chain. It deals with all parts of the signal acquisition and the description of the signals used for the disaggregation. The conditioning of the signals for the following event detection and classification is also an essential part of this chapter.

This chapter describes the signals that can be used for the disaggregation. For this a system model was developed. This shows the appliances in the load profile of residential buildings. The household is part of the Smart Grid. NALM means, that the signals are measured at a central point. In our case at the household entry point. From the physical variables $u(t)$ and $i(t)$ further signals can be calculated. The derivation of the power signals P , Q and S were described. A normalization of these signals has also been described by the model. This was necessary to eliminate the influence of the grid resistance.

Disturbances which occur due to the measurement were described. Firstly, the influence of noise has been described on the measurements. By the measurement, white noise is superimposed additively. Another influencing factor is the digitization of analog continuous-time signal. At the AD conversion quantization errors occur. Both factors determine the minimum resolution at which appliances can still be detected.

Since the beginning of the work were no measurements available. For this reason own measurements were performed. These measurements are later used for the development and statistical evaluation of disaggregation algorithms. I have defined requirements, which should fulfill the measurement system. Based on these requirements, the measurement system has been developed. It was followed by the statistical analysis of measurement systems.

Following the development of the measurement system own data were recorded. For this purpose, several individual appliances in residential buildings were measured. Also pattern measurements were simulated. With the measurements the behavior of the individual appliances in the load profile were analyzed. Different criteria (such as transient, harmonic, ...) were analyzed. Also could be detected by the plurality of appliances, whether they did not differ by manufacturer.

At the end of the chapter a filter has been presented. With that the output signals were pre-filtered. This filter should minimize peaks, which are caused by distortions in the grid. But the waveform should not be distorted. For filtering the signal, a median filter was used. The filter was adjusted to the signal. The ideal filter length was determined. The filter was simulated and verified by real measurements.

The measurements in the database are from residential buildings. However NALM can also be useful in the industrial environment. For this purpose, additional appliances must be measured which arise in the industry. Since the measuring system can measure currents up to 63 A further measurements can be performed up to this value.

5 Event Detection and Detector

Contents

5.1	Introduction	60
5.2	Problem Statement	61
5.3	Methods	64
5.4	Event Detection Algorithm	68
5.4.1	General Structure of the Event Detector	68
5.4.2	High-Pass Filter for the Event Detection	69
5.4.2.1	Approach of Hart	69
5.4.2.2	High-Pass Filters from Image Processing	70
5.4.2.3	The Gradient Operators	71
5.4.2.4	Laplace-Operator	73
5.4.2.5	Simulation of the High-Pass Filters	74
5.4.3	Complex Filter Structure	76
5.4.3.1	Approach of Canny	76
5.4.3.2	Improvement of Canny Filter by Perona	78
5.4.3.3	Simulation of Peronas Filter	81
5.4.4	Short Time Fourier Transformation	83
5.4.4.1	Simulation of the STFT	84
5.4.5	Classification Function	87
5.4.5.1	Threshold for the Classifier	88
5.4.5.2	Simulation of the Classifier	89
5.5	Results for the Event Detection	90
5.5.1	Experimental Procedure	90
5.5.2	Best Signal for the Event Detection	91
5.5.3	Results of the Event Detection Methods	91
5.6	Conclusion	94

In this chapter, the event detection is described for the disaggregation as a preprocessing for classification. An overview will be presented of current methods of event detection in digital signal processing. Algorithms are then derived which can be used for the event detection in NALM. The functionality of the algorithms is simulated. At the end of the chapter the algorithms are tested for usability and the quality of the detection rate is discussed. Therefor the results of the event detection for real measurements are presented.

5.1 Introduction

Events are changes in the mean level of signal curves or time series. It will be distinguish between different kinds of events, continuous and abrupt. In the English literature the detection of events in multidimensional signals is called *Edge Detection*. Edge detection is found e.g. in image processing for the segmentation of elements. It exist different mathematical methods to identify discontinuities in the image. Edge Detection is also found in computer and machine vision. The detection of events in one dimensional signal is called *Step Detection* (Sowa et al., 2005) (don't confuse with walk detection). The signals used in this work for the disaggregation have as parameters only the time t . So they are one-dimensional. Therefore, we should speak of step detection. But in the methods of NALM it has established the name *Event Detection* (ED) (Najmeddine et al., 2008). So in this work we speak from Event Detection to identify discontinuities in the measured signals for the disaggregation.

The Event Detection (ED) is necessary, because a real time system should be developed. In a real time system for NALM, the signals $u(t)$ and $i(t)$ are measured continuously (cf. chapter 4). Therefore, the disaggregation must be performed online. In the state of art, different methods for a continuously classification were presented. For these events detection is not necessary. The event itself is a feature in the feature space for classification. But, a continuously classification of the signals demands in a higher calculation effort and the requirements to the hardware are increasing. The result is, that the costs of the system will be increase and the acceptance of such a system in the population will be decrease. In current smart meter systems it is not possible to make a continuously classification. For these reasons, the disaggregation is done offline for these methods which using only a classification. But in this work, the use of disaggregation on current smart meter systems should be investigated. Therefore, a pre-processing and a reduction in the measurement data have to be performed. For this reason for a real time system it is necessary to recognize the switching events continuously. Through the detections of the switching on events it is possible to classify directly at the switching on point. As a result the calculation effort for the classification is decreasing.

The event detection is a main part of the disaggregation system presented in this work. After the measurement and conditioning the signals (cf. 4) the second main block of the disaggregation-chain follows. The block is called Event Detector, consists of the event detection algorithms and is shown in Fig. 5.1 as part of the whole system (Benyoucef et al., 2011b). The input signals of the event detector are the time series of the power signals (P, Q, S) described in chapter 4. This allows a temporal resolution

of the signal from 20 ms. The switching on- and off-cycles of appliances in the overall load profile results in an abrupt change in the power signal. These changes must be detected by the event detector. The output of the event detector provides the time-stamps of the switching cycles to the following classifier. The classifier then leads on to the detected point, the classification of the appliances.

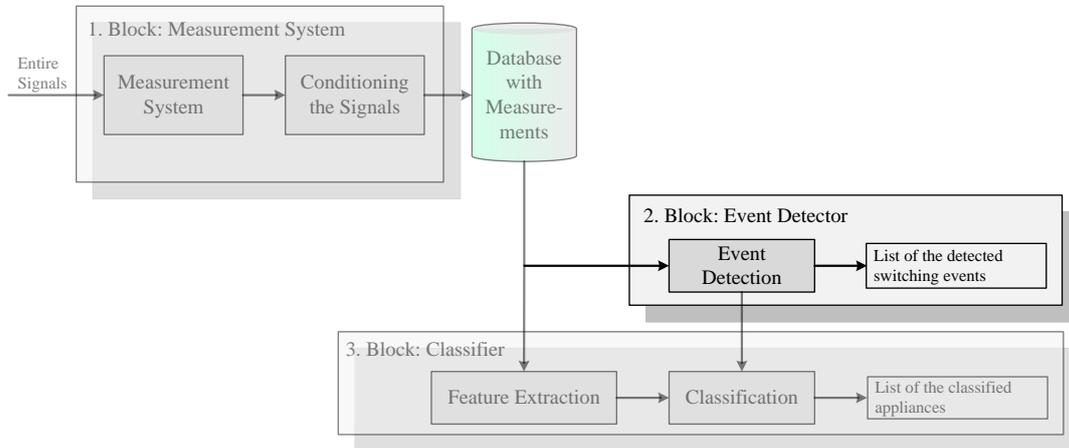


Figure 5.1 – Disaggregation system: the event detector.

An example of a real measurement is shown in Fig. 5.2. General in NALM it must be distinguish between 2 kinds of events:

- ▶ switching on events, where the appliance is switched on. This event is important for the classification of the appliance, and
- ▶ switching off events, where the appliance is switched off. The detection of the switching off events is necessary for the tracking and the calculation of the power consumption of the appliances. After a switching on event is detected and the appliance is classified, the power is integrated over the time until the switching-off event. Hence the consumed energy of the appliance is calculated. In a bill, these calculated energy consumption can be shown the resident of the household. This can be used to calculate the individual energy consumption of each appliance. Analogous to the switching on events, it is possible to detect some error events. Nevertheless, a switching off event is always a direct cut in the energy consumption.

The tracking of the energy consumption is not part of this work. For this reason, it will not be strongly distinguish between these two types of events for the derivation of the algorithms. Where it is necessary to distinguish between the two types of events it will be done explicit.

5.2 Problem Statement

As just described, there are various types of events. The signal can be superposed with disturbances. This results in errors in event detection. The disturbances have different physical properties. Firstly, they are created by noise, which are superimposed on the

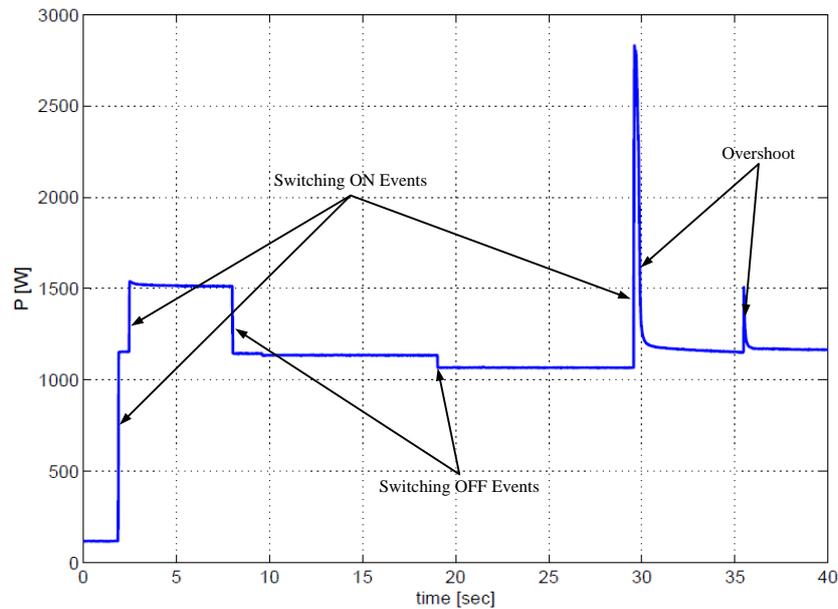


Figure 5.2 – In NALM two kinds of events exist, switching-on and switching-off events. The figure shows an example of a real measurement. In this case a water-heater, a hair-dryer, a refrigerator and a radio were switched on.

signal in the measurement by the measuring system (cf. chapter 4). Another type of interference caused by the different physical behavior of the electronic appliances. Similarly, disturbances can be injected by the grid. In order to develop an event detector, which is robust against disturbances, it is necessary to know them. By research in the literature (Canny, 1986) and own preliminary investigations and measurements, the following basic types of events with disturbances were identified. In Figure 5.3 these are represented graphically. They show

- a) an idealized on/ off-event of an appliance. Here the steady state is reached after a sample ($f_s = 50$ Hz), and no fault can be detected. In this scenario the ED should recognize ideally only a switching on event (at t_1) and a switching off event (at t_2).
- b) a ramp-like profile. However, the measurements and preliminary studies have shown, that there are no devices that require more than 3 samples (similar to 60 ms cf. chapter 4) to detect a significant increase in the power. Ramps with a lower gradient occur in more complex appliances such automates like TV. A detection of switching as this is undesirable due to the later computational complexity in classification. The gradient of the ramp may vary depending on the appliance.
- c) an oscillatory disturbance that occurs through overshoots. After a switch-on event an overshoot can cause. These can be detected by the ED as turn on or off multiple appliances. The overshoots have a physical cause, the structure of the consumer. E.g. it requires a time until the internal power storage of the power supplies of the appliances are recharged. An overshoot does not occur for switching off events. Because appliances are separated hard from the grid.

- d) disturbances caused by peaks. Individual measurement values are detected as outliers. This is not to detect as switching-on or switching-off event should be specified for the ED. By the measurements, it was found that such disturbances can be caused by the grid. A possibly pre-filtering of the signal minimizes these problems. One method, the median filter was described in the past chapter 4.7.
- e) interference from noise. White noise is caused by the measurement (cf. chapter 4). Is the noise amplitude greater as the power amplitude of the appliance, the switching events of the appliances are disappear in the noise.

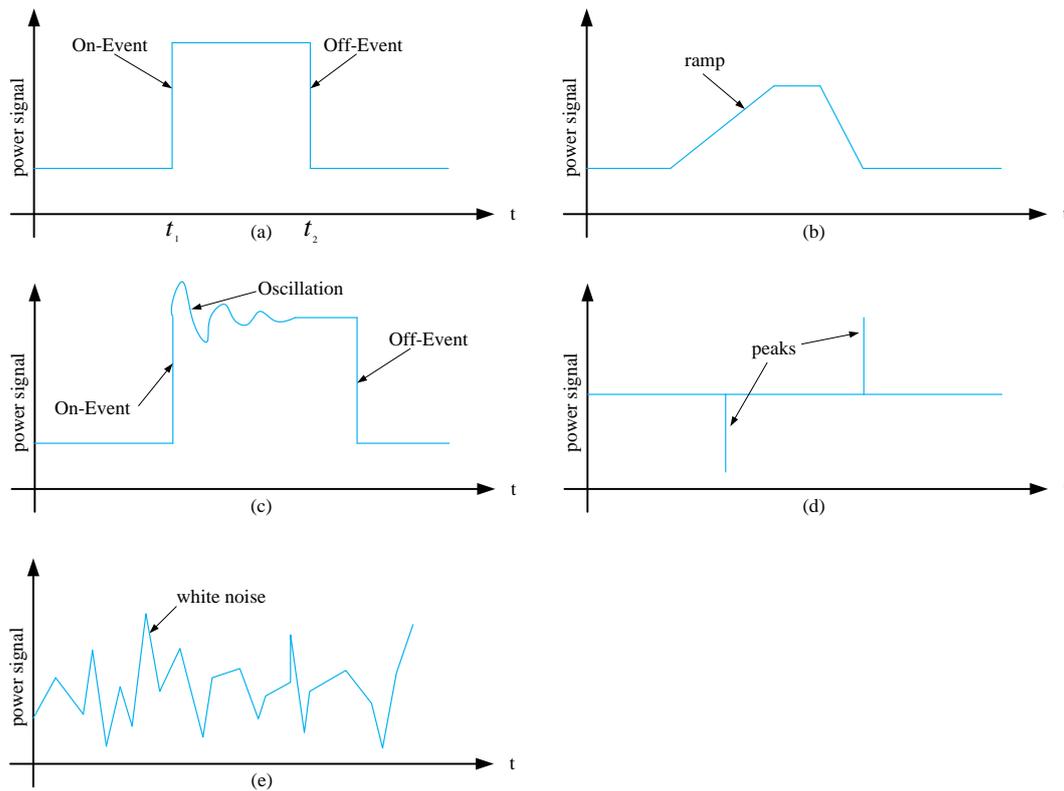


Figure 5.3 – Possible switching events with and without disturbances.

Based on those 5 types of possible states the following three assumptions are defined for the development of an event detector:

1. Find the switching on and off events of appliances with a power consumption between 5 W and 3680 W.
2. Detection of abrupt changes in the power signal. From the measurements we know, that the appliances needs between one and three samples to reach their maximum of energy. Slower gradients with $\Delta P < 5$ W should not be detected as an event.
3. The two disturbances from Fig. 5.3d and 5.3e should be minimized.
 - Case e, peaks from the grid are eliminated with the median filter (cf. chapter 4.7).

- Distortions from noise (AWGN cf. 4.3.1) should be reduced by the event detector.

5.3 Methods

The approaches of the event detection are divided in the literature (Basseville and Nikiforov, 1993) into two groups. It will be distinguish between *Online change detection* methods and *Offline change detection* methods. When the data arrives and when the event detection must be performed as, then online algorithms are usually used (Basseville and Nikiforov, 1993). Then it becomes a special case of sequential analysis. In the other case, offline algorithms are applied to the data potentially long after it has been received. Most offline algorithms for step detection in digital data can be categorized as top-down, bottom-up, sliding window, or global methods. The methods presented here are from the field of online change detection method. They were selected for subsequent implementing to a real-time system. These methods have the assumption that the measured signal is a random process. The methods come from different areas of electrical engineering. During their research, many methods have been found in the field of image processing. The applicability to the present problem is discussed here. Furthermore, methods are presented which are already used in NALM systems.

„Nonintrusive Appliance Load Monitoring (NALM)“ Approach of George W. Hart 1992 (Hart, 1992) As in the state of art 3.1.1 described, Hart works on the disaggregation of appliances in the load profile of residential buildings. For the event detection, he uses the power signals. He performs a normalization to the mains voltage. The normalized power is transferred to the ED. It detects the time and the amplitude value of the event. To calculate the power leap he waits some time until the steady state is reached. This leads to peaks and overshoot occurring as noise are not detected. The tolerance, in which the signal may oscillate in steady state is about 15 W (or 15VAR). To minimize noise, he integrates the values at steady state. He uses this mid value to perform a differentiation of the signal. These delta values it leads to a classification function with threshold value. Depending on the set threshold, he classified the calculated value as an event. The threshold value Hart places either on the level of power consumption of the appliances to be discovered. In his algorithm described for large appliances. The temporal resolution of the switching events is a few seconds until to 15 min.

This method can detect events well if the appliances have few states. It was designed for appliances who have a relatively high energy consumption (greater 35 W). A detailed description of the procedure and the parameter set used could not be determined because the detailed description of (Hart, 1985) at MIT was not present. Switching times of less than 1 s are not detected by the small sample rate. This method provides an easy way to detect events by differentiation of the signal. That is why it was picked as the first method to be evaluated. The method was developed for a specific group of appliances. Therein lies its weakness. In this thesis, appliances should also be detected with low energy consumption. Therefore, this method needs to be optimized.

However, significant improvements are still to be included. The improved algorithm is described in 5.4.2.

„Using a Pattern Recognition Approach to Disaggregate the Total Electricity Consumption in a House into the Major End-Uses“ Approach of L. Farinaccio and R. Zmeureanu (Farinaccio and Zmeureanu, 1999) The approach of Farinaccio and Zmeureanu based on two approaches. The first one is the approach already presented by Hart (Hart, 1992). The second approach is the Heuristic End-Use Load Profiler, developed by Quantum Consulting in California (Powers et al., 1991). Like Hart, training data are first taken to obtain the signature of every electrical appliance. They assume that no appliances at the same time run. They set for each appliance a proprietary algorithm, whose goal is to identify certain characteristics reproducibly again. They distinguish between the on-, run-up and off-events. It is not directly performed an ED with subsequent classification of the appliance. Further, the signal is continuously compared with the individual characteristic curves, which corresponds to the principle of a correlation. If the consumption changes within a year, some of the algorithms must therefore be adjusted again.

The method is based partly on the approach of Hart and thus has its strengths and weaknesses. Because of the high threshold, it has a low error rate. At the same time only appliances with large consumption can be detected. Even with a superposition of multiple appliances no new cycles can be detected. This approach is not suitable due to its many requirements for a practically oriented ED.

„Cumulative Sum Charts and Charting for Quality Improvement.“ Approach of Douglas M. Hawkins and David H. Howell (Hawkins and Howell, 1998) The cumulative sum (CUSUM) examined changes in a sequential measurement or data sequence. It is a method of analysis of the statistical process and quality control. It has its origins in the prediction of stock prices. In 1998, Howell and Hawkins described this method. It is not just the pure cumulative sum of the data values, but the cumulative sum of the differences between the data values x and the predefined data values w . Formally, this means $S_n = \max(0, S_{n-1} + x_n - w_n)$, n is the current measured value. To see now whether it is a rising, falling or unchanged function, a so-called V-mask is now drawn. Initially, two not very distant points from each other are determined. This means in particular that set a point $\{n, S_n - h\}$ first. From this then a line with a certain slope $k = \frac{1}{2}\Delta$ is drawn. Thereafter, the position of the second predetermined point is determined. This is below the line, this indicates an increasing change. For the descending change starting at the point $\{n, S_n + h\}$ and $-k$. This produces the eponymous V shape.

According to the investigations of (Hawkins and Howell, 1998) this method is well suited to detect monotonic inclines and declines. It is also robust against white noise. Major drawback of this method is the detection of events or steep inclines. Similarly, it is poorly suited for the detection of complicated signal patterns. This method is more suitable to determine long-term changes in the signal. This approach violates the second condition. It is not possible to detect abrupt changes in the signal. Hence this method was not pursued due to its weaknesses.

„A Computational Approach to Edge Detection,“ Approach of Canny 1986 (Canny, 1986): Another approach is the use of filters for the detection of events. A basic example in this area is the approach of John Canny. In 1986, he established three general sets of correct recognition, which are mainly used in image processing today. They are as follows:

1. Good detection of an event. This is equivalent to maximizing the signal-to-noise ratio.
2. Good localization. The points which are detected as an event should be close to the center of the real event.
3. One event should be recognized only once to avoid errors.

According to Canny events and their location should be easily identified if you stick to these principles. Also noise or shades (in pictures) should not be falsely detected as information. In summary, Canny (Canny, 1986) rated his method good for the detection of events in two-dimensional signals. It has a low error rate in terms of not detected events. Due to these strengths, there is one widespread method of event detection in image processing. His weaknesses are, according to (Canny, 1986) in the noise sensitivity. This approach violates the third condition. It has a poor noise cancellation. Another problem is the reliable detection of features that do not consist of simple steps, examples are ramps, roofs or peaks. This approach is promising in its fundamentals but contains some weaknesses. Later research groups took up this approach and improved it.

„On Optimal Linear Filtering for Edge Detection“ Approach of Didier Demigny (Demigny, 2002) The approach of Demigny is very similar to the approach of Canny (Canny, 1986). His work is based on Canny’s three principles of good detection, good localization and detection of an event only once. The difference to Canny’s work is that he uses not a single but various stages function as the input pulses of different widths d . Thus, the formulas presented by Canny are change. The next point Demigny has an optimal filter defined. Instead, he has created a table with some coefficients to help with the selection. But this is not discussed in detail.

The approach is strongly applied to those of Canny and thus also its advantages and disadvantages. The selection of the pulse as a step function means that this method is rather designed for 2-dimensional signals (in that case images). Since a filter based on algorithm already been found by the method of Canny, this method was not pursued further because of the weaknesses to condition 3.

„Optimal Edge-Based Shape Detection.,, Approach of Hankyu Moon, Rama Chellappa and Azriel Rosenfeld; 2002 The group of Moon (Moon et al., 2002) has been viewed several approaches and has come to the target that a differential calculation should provide optimal results for the event detection with noise suppression. For this purpose, they used the DODE function (Derivative Of the Double Exponential, also known as Laplace distribution). For the input function, a similar function of Canny approach (Canny, 1986) was defined. To this, with the amplitude α for $x \geq 0$ a

noise term was also added. It was AWGN. Next, the addition, of the root mean square of the difference between input and output signal, and the output noise response is minimized. After the derivation of the spectral densities (Moon et al., 2002) this leads to the Wiener filter. This is considered herein to optimal filter and used in the structure of the Canny approach (Canny, 1986). The transformation must be adapted to each object to be recognized.

Since this approach is based on the principles of Canny it is also optimized for applications in image processing. Therefore, it aims more on 2-dimensional signals. The major disadvantage of this algorithm is the necessary knowledge of the object to be detected (Moon et al., 2002). Similarly, the threshold of the classification function must be adjusted each time new. These a-priori information unusable this approach to be developed here for the ED. This approach violates the first condition.

„Detecting and Localizing Edges Composed of Steps, Peaks and Roofs “ Approach of Perona and Malik 1990 (Perona and Malik, 1990): This approach is a further development of the previously proposed approach of Canny. Due to shading or noise occurs in Cannys approach still mistakes. Similarly, all other functions except stages poorly recognized. For these reasons, Perona and Malik have improved Cannys approach in 1990. They have retained the principles, but they have, e.g. making changes in the filter selection. The following results are presented briefly. They are using a square filter structure and not a one filter structure. They also used two different filters. A filter is responsible for the odd part of the input signal. This reacts on stairs or rectangular functions. The other filter responds to the even part. These are, for example, peaks or ramp functions. After several attempts, Perona and Malik came to the realization that for the odd filter a twice derived Gaussian window, and for the even filter the Hilbert transform of this should be used. This filter structures have been included in the base formulas of Canny, with some minor changes. This changes are described and discussed in detail in chapter 5.4.3.1.

Malik and Perona evaluate their approach as follows (Perona and Malik, 1990). To strengthen the approach of Canny that events are well detected, added even more benefits. The noise sensitivity decreases as compared to Canny. As a result, the error rate decreases. This extended approach can detect ramps and also peaks. This approach looks promising. Initial evaluations in the laboratory have shown that the algorithm detects events well. This approach fulfills the predetermined assumptions 1-3. It is therefore promising for event detection. For these reasons, it is described and analyzed in chapter 5.4.3.1 with other methods for their usability as an ED algorithm for NALM.

STFT The methods shown previously, use for event detection the time series. Another possibility of the detection of events can be the analysis of the signal in the frequency domain. The spectral properties of the signal can be calculated using the Fourier transform (Kammeyer and Kroschel, 2012). Analyzing a time-varying signal may be performed with an enlargement of the FT, the so called STFT. This method can be used to detect changes in signals (Moukadem et al., 2013).

The methods presented here are concerned with the detection of events in the signal paths. The advantages and disadvantages of the method are briefly described here. The method will be divided into 4 groups. The first group conducts research in the area NALM. The first approach was developed by Hart and by other groups used such as Farinaccio for event detection. The second approach used the CUSUM method. The third group used some methods for event detection from the field of image processing. Based on the approach of Canny other researchers work in this field. Perona, Malik, Demingy and Moon have the method further improved and adapted to their questions. The last group deals with the analyze of the signal in frequency domain. Therfor the STFT can be used.

In the following three approaches are analyzed more in detail. They will be improved for our problem, to detect events for the disaggregation. At first an improvement of the approach of Hart will be developed. It based on a filter structure using high-pass filter. The second approach used also a filter but a more complex one. It based on the approach of Perona. The last method to approach events is based on the STFT, to find frequency components for the detecting of events.

5.4 Event Detection Algorithm

In this section the derivation of the algorithms for the event detection is described. First, an extended form of the Hart approach is described. This performs a differentiation of the signal. The implementation into a filter structure is described. Then the approach of Perona is derived. This uses also a filter structure for the detection of events but with a more complex filter function. Finally, an approach is described using the STFT for a time frequency analyze of the signals. The detection rates of the three approaches are simulated with idealized signals. At the end, the quality criteria of the algorithms are presented and discussed by using real measurements.

5.4.1 General Structure of the Event Detector

The structure of an event detector is in the most cases the same. It can be divided into two blocks which consists of the decision function and the classification function. As described in chapter 2.3 the decision function assigns each input to a corresponding numerical value as output. The classification function then assigns the decision function value to a class label. In our case it is a binary classification problem

- ▶ An event is detected or
- ▶ an event is not detected.

In figure 5.4 the block diagram of the processing chain is given. The input signals are the conditioned output signals P, Q and S. For the decision function the three described approaches are used. The classification function is a threshold decision maker. This then places the decision function value for the three approaches the label *event* or *no event* too. The output of the event detector is the trigger for the classifier.

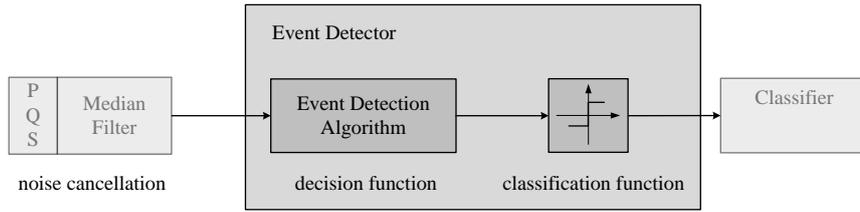


Figure 5.4 – Block diagram of the event detector as part of the disaggregation chain.

For the description of the algorithm the active power P is used. But it is also possible to use the reactive or apparent power or a combination of them for the detection of events. The best signal in sense of optimizing the detection rates will show in section 5.5.2.

5.4.2 High-Pass Filter for the Event Detection

Through the frequency analysis it is known that rapid changes in the signal contains of high-frequency components. Slow changes have correspondingly low frequencies. A high-pass filtering may thus emphasize abrupt changes in the signal and allow the detection of event.

The first approaches of NALM use a differentiation of the signal for event detection. This has been applied in several approaches and should be further analyzed here.

5.4.2.1 Approach of Hart

The first algorithm for NALM based on the approach of Hart (Hart, 1992). He used a differentiation of the signal to detect events in the load profile. The algorithm presented here calculates delta values of the signals power for the event detection. The values of the differentiation are calculated for time continuous signals as

$$\Delta p(t) = \frac{p(t) - p(t - \Delta t)}{\Delta t}. \quad (5.1)$$

$\Delta p(t)$ is the current delta value of the power at time t . $p(t)$ is the (aggregate) current power consumption of the appliances. $p(t - \Delta t)$ is the power value in the past where Δt is a user-defined time period.

Other approaches wait with the differentiation of the signal until the appliance has swing out (Hart, 1992). The steady state, can often take several seconds, as is known from the preliminary studies. The approach presented here should be optimized so that an event is detected immediately as possible. In the optimal case, Δt should be chosen as small as possible. Under the consideration of the possible waveforms from Fig. 5.3 only case b) is critical. A ramp where the gradient is small would not be recognized properly. However, this case can be excluded from the preliminary studies described above. All other cases would be detected by the sliding window as an event. For a sampled signal the time of a continuous signal $p(t)$ becomes $p(n \cdot T_s)$ with $t = n \cdot T_s$. $T_s = \frac{1}{f_s}$ is the period between two measured samples, f_s is the sampling frequency.

By normalizing T_s to the sampling frequency we get $P(n)$ as the new sampled signal. Under the assumption, that the time between the two measured values are minimized, T becomes 1. Then eq. (5.1) becomes

$$\Delta P(n) = P(n) - P(n - 1). \quad (5.2)$$

This equation was implemented by Hart. The main disadvantage of Hart's approach is, that disturbances which have a greater amplitude than $\Delta P(n)$ are also detected as an event. So this approach has a poor noise cancellation. This approach is the base for the following idea to use a high-pass filter for the event detection.

5.4.2.2 High-Pass Filters from Image Processing

The differentiation of the signal is similar like high-pass filtering. To reduce the faults from noise, other high-pass (HP) filters with different structures should be analyzed. In image processing it gives different operators to detect events in one or two dimensional signals. Every filter consist of a kernel which is called mask. The outputs of these filters are the convolution integral of the mask f and the input signal $P(t)$. For a better comparison of the different methods, the nomenclature of the Δ is also used as output of this filter. The output of the HP filter ΔP_{HP} is calculated as

$$\Delta P_{HP}(t) = (f * P)(t) = \int f(\tau) \cdot P(t - \tau) d\tau. \quad (5.3)$$

At the beginning of the development the best known operators in the field of image processing were investigated. Depending on the application, they can be applied to one or two dimensional signals. In the following these processes are briefly described.

- ▶ The *Sobel-Operator* is a simple edge detection filter. It is frequently used in image processing. It is used there as the convolution operator in the so-called *Sobel-Algorithm*. This computes the first derivative of the pixel brightness values. At the same time the filter smoothes orthogonal to the direction of derivation. The algorithm uses a convolution by means of a 3x3 matrix (convolution matrix) for generating a gradient signal from the original signal. This high frequencies in the signal are shown as high output values. The areas of greatest intensity are where the brightness of the original image has the most changes and thus represents the largest edges. Therefore, after the convolution with the Sobel operator is a threshold function applied.
- ▶ The *Roberts-Operator* is a simple edge detection algorithm of image processing. It is one of the oldest operators. The operator was presented by Lawrence Roberts (Roberts, 1963) in 1963. It calculates the difference between the lying crosswise pixels. Therefore, this operator is also referred to as Roberts cross operator. The output signal is to be calculated quickly and easily. However, the operator is poorly suited for noisy images.

- The *Laplace-Filter* is a filter for edge detection. It approximates the Laplacian operator. This operator based on the second derivation.

All these operators differ in their filter cores. For the problem described here a linear one-dimensional high-pass filter is searched. The first two operators are derived by the so-called gradient method. The last one is based on the derivation of the Laplace-operator. Except for the Laplace filter all approaches have been developed for two-dimensional signals. Therefore for the one-dimensional signal own filter masks had to be derived.

5.4.2.3 The Gradient Operators

In image processing two-dimensional filters are used to detect edges in images. For the edge detection, the signal can be considered as a measured value function $p(n,m)$. Then edges are steep gradients in p . Thus the derivative of the function is the key for edge detection. The gradient of a continuous, differentiable, two-dimensional function $p(n,m)$ is defined as:

$$\Delta p(n,m) = \frac{\partial}{\partial n} [p(n,m)] e_n + \frac{\partial}{\partial m} [p(n,m)] e_m \quad (5.4)$$

n and m are the continuous space coordinate. e_n and e_m are the corresponding unit vectors. We obtain a vector of magnitude and direction. This indicates the maximum first derivative and its direction in the nm – *Level*. Based on this definition, we are now looking for approximations. This should specify the derivation of the continuous function $p(n,m)$ by local differential developments of the corresponding stationary discrete function. Here the measured values of the source signal are weighted within a local window and then used for gradient formation. Dependent on the selected operation the result is assigned to a certain measuring point of the local window. So it is an operation that generates the results step by step using the measured values.

We first choose the smallest possible window size 2x2. This provides an easy way of linking the measurement point. It is called the *left-side differentiation*.

$$\Delta_n [p(n,m)] = p(n,m) - p(n-1,m) \quad (5.5)$$

$$\Delta_m [p(n,m)] = p(n,m) - p(n,m-1) \quad (5.6)$$

The approximation of the derivative operator on a discrete signal leads to the difference operator. This neutralized constant signal parts and highlights measured value differences. For the calculation of the gradient for a two-dimensional signal the two results must be added together.

Up here the derivation of a gradient operator for a two-dimensional signal was observed. Now the special case of a gradient for a one-dimensional signal is to be considered. This simplifies eq. (5.4). For the derivation of the one-dimensional case in n -direction

eq. (5.4) becomes

$$\Delta p(n) = \frac{\partial}{\partial n} [p(n)] e_n \quad (5.7)$$

Eq. (5.7) should be used in the following to derive the first two operators. The approximation of the derivative operator on the discrete-time signal is given by

$$\Delta_n [p(n)] = p(n) - p(n - 1). \quad (5.8)$$

The filter mask for this operator is

$$f = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (1 \quad -1)^T. \quad (5.9)$$

This forms the smallest possible mask length. It is the easiest way of linking measurements and is referred to as the left-side difference. Eq. (5.8) refers mainly to the difference equation of Hart (cf. eq. 5.2) and thus also has its disadvantages. An important is that the noise is not minimized. But there occurs still a further effect. By the left-side difference, we see that the result includes the actual measurement value $p(n)$ and the last measured value $p(n - 1)$. This is but then again assigned the time value n , as a measured value in between does not exist. The gradient is thus shifted to the measurement signal. In this case by $n/2$. This effect does not interfere at first, but must be taken into account when determining the switching time of the event.

To avoid this effect it can be defined a differential operator, which includes a mask with three values. These operators are called central differences. Let us consider directly the one-dimensional case. The difference equation is

$$\Delta_{2n} [p(n)] = p(n + 1) - p(n - 1) \quad (5.10)$$

The operator for this equation does not include the actual value of n . It includes, in contrast to the first definition of each direction a distance from one measurement value. The response of the operators Δ_{2n} on an ideal event has also a width of two values. The localization accuracy is therefore worse than by the two value mask. The filter mask for this gradient operator has three values

$$f_{\Delta_{2n}} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = (1 \quad 0 \quad -1)^T. \quad (5.11)$$

This one dimensional mask is alike to the two dimensional Sobel Mask. Moreover, the mask size of 3 values improves the effect of reducing noise.

From the measurement it is known, that appliances need 1 until 3 samples to reach the steady state. To handle this eq. (5.10) is extended over one value.

$$\Delta_{3n} [p(n)] = p(n + 1) - p(n - 2) \quad (5.12)$$

We get a left side difference. Here the value from two times in the past and the next

value from n are used to calculate the gradient. The mask from eq. (5.12) is

$$f_{\Delta_{3n}} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = (1 \ 0 \ 0 \ -1)^T \quad (5.13)$$

All operators discussed so far are based on differences of weighted averages. Subsequently, another definition for the derivation of a mask will be described.

5.4.2.4 Laplace-Operator

The operators described above are discrete local operators. These are defined on the basis of the continuous gradient. Now a mask should be determined more in detail based on the continuous Laplace operator. For a two dimensional signal it is defined as

$$\nabla^2 p(n,m) = \frac{\partial^2}{\partial n^2} [p(n,m)] + \frac{\partial^2}{\partial m^2} [p(n,m)]. \quad (5.14)$$

The operator is used in the so-called Laplacian filter, which is widely used in image processing. The discrete Laplace operators react very strongly to corners, lines, line ends and isolated points in the signal. Thus condition 2 would be achieved. However, this behavior can be unsatisfactory in a noisy signal which would be against the assumption 3. We will see this in the simulations later.

Eq. (5.14) is defined for a two dimensional signal. For a time discrete one dimensional signal the first derivative of eq. (5.14) is for two measured values

$$\Delta_{n1} = p(n) - p(n-1) \quad (5.15)$$

$$\Delta_{n2} = p(n+1) - p(n) \quad (5.16)$$

where Δ_{n1} is the left-side difference and Δ_{n2} is the right-side difference of measured point n . For the twice derivation, the two gradient values eq. (5.15) and eq. (5.16) are differentiated again. The gradient at measured point n is now

$$\begin{aligned} \Delta_{n1}^2 &= \Delta_{n2} - \Delta_{n1} = p(n+1) - p(n) - [p(n) - p(n-1)] \\ \Delta_{n1}^2 &= p(n+1) - 2 \cdot p(n) + p(n-1) \end{aligned} \quad (5.17)$$

The mask for the discrete one-dimensional Laplace operator can be derived from eq. (5.17). The mask is

$$f_{Lap} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = (1 \ -2 \ 1)^T. \quad (5.18)$$

There are 3 HP filters were presented. The masks of the filters are shown in Fig. 5.5. Two masks based on the gradient operator. $f_{\Delta_{2n}}$ and $f_{\Delta_{3n}}$ point-symmetric. The mask based on the Laplacian Operator f_{Lap} is axis-symmetric. We will see later, that this symmetrical relationship has an influence for the differentiation between a switching on and off event.

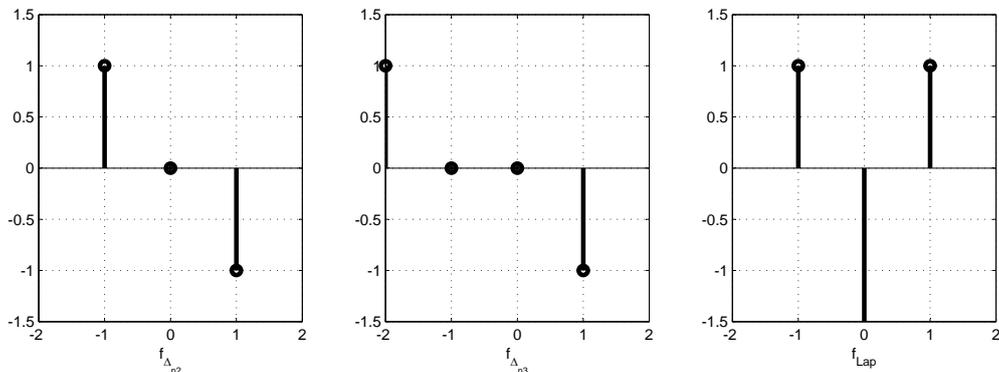


Figure 5.5 – Masks of the three high-pass filters. The masks $f_{\Delta_{2n}}$ and $f_{\Delta_{3n}}$ based on the gradient operator. These are point-symmetric. The laplace mask f_{Lap} based on the Laplace Operator. This mask is axis-symmetric.

5.4.2.5 Simulation of the High-Pass Filters

The masks previously defined based on gradient and Laplacian operators. As described, these have a better noise reduction compared to the approach of Hart. In the following simulations are carried out, which should analyze the theoretical foundations. Based on the system characteristics the effect of superimposed noise on the detection rate has been investigated for the individual masks.

The noise is AWG (cf. chap 4.3.1). The noise power of n is calculated as a function of the standard deviation σ

$$n_{dB} = 10 \cdot \log_{10} \left(\frac{\sigma^2}{1 \text{ W}} \right). \quad (5.19)$$

The noise power is given logarithmic. It refers to 1 W. The simulations were done with different noise powers. From eq. (5.19) σ can be derived.

$$\sigma = \sqrt{10 \frac{n_{dB}}{10}} \quad (5.20)$$

Dependent from the signal amplitude and the intensity of the noise the SNR (cf. chapter 4.3.2) can be calculated. For a noise power of $n_{dB} = -0.8505$ the SNR is

calculated with eq. (5.20) and eq. (4.17) as

$$\sigma = \sqrt{10^{\frac{n_{dB}}{10}}} = \sqrt{10^{\frac{-0,8505}{10}}} = 0,9067$$

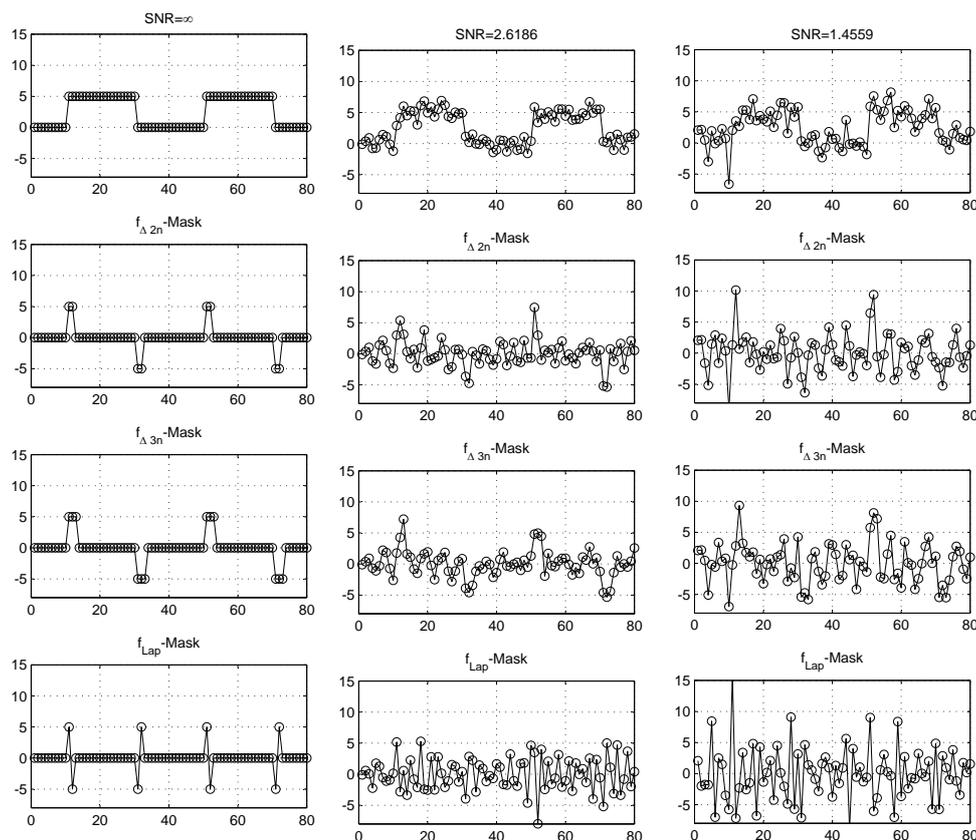
$$SNR = 10 \cdot \log_{10} \left(\frac{A}{\sigma} \right) = 10 \cdot \log_{10} \left(\frac{2.5}{0,9067} \right) = 2,75$$

The results of the simulation with different noise levels are shown in Fig. 5.6. In the first row the input signal of the filters is shown. The signal has a amplitude of 5 W. To cycles with 4 evens are shown. The input signal is superimposed by AWGN. In the first column an idealized signal is shown. Here the $SNR \rightarrow \infty$. In the second column, the noise power was increased. The SNR is still 2.7572. The events are still visible to the eye. In the third column σ is further increased. It is increasingly difficult to identify the individual events yet. The SNR is still 1.5. In the second line the output of the filter with the filter mask according to eq. (5.11) is to see. The third line shows the output of the filter with the mask of eq (5.13). In the last line the output of the Laplacian filter according to equation (5.18) is shown. The first and second filters is detected the events for the SNR of 2.7. The Laplace filter here has been significant problem. For a SNR of 1.5, all three filters have a poor detection rate. Based on the simulated evaluation shows that the mask $f_{\Delta_{3n}}$ has best properties in terms of the defined assumptions 1 and 3.

All three methods can also distinguish between an on- and off-event. The two gradient methods are point-symmetric. These have for a rising event a positive and for a falling event a negative output value of the filter. The filter with the Laplace operator is axis-symmetric. It obtained positive and negative values for both rising and falling events. However, here the sign is reversed. With a corresponding classification function it is possible for all three methods to differentiate between an on and off events.

In the previous simulation, the gradient of the events was $\Delta = \frac{5}{1}$. That means, an appliance with 5 W power consumption achieved the steady state in one sample. However in the second assumption other gradients were determined. This assumes that there are appliances which need 2 or 3 samples to obtain the steady state. Here the gradient is then $\Delta = \frac{5}{2}$ and $\Delta = \frac{5}{3}$. The simulation in Fig. 5.7 analyzed which mask detects this gradient best. In the first row the idealized signal is shown. Now with different gradients for the switching on event. In the rows 2 until 4 the outputs of the three filters are shown. The smaller the gradient of the switching on event is, the low the output amplitudes of the individual filters. Only the filter with $f_{\Delta_{3n}}$ mask reached at the lowest gradient, the maximum output amplitude 5 W. This shows again that the $f_{\Delta_{3n}}$ mask has the best properties in terms of assumption 2.

The simulated filters are suitable to detect events. It has been found that events are also detected when the signal is superposed with noise. Because of the short filter length they are well suited for implementation in real-time system. However, in the simulations also be seen that when the Signal to Noise Ratio (SNR) is too small a relatively high number of False Positives (FP) are detected. In the following section a further filter will be presented. This promises, due to a complex structure of the filter function, a further reduction of the FP for the same detection rate as the HP filters.



(a) An idealized waveform (b) The signal with AWGN and $SNR \rightarrow \infty$ (c) The signal with AWGN and $SNR = 2,75$ and $SNR = 1,54$

Figure 5.6 – Simulation of the high-pass filter. The first row presents the signal with AWGN. The second row the output of the filter with $f_{\Delta 2n}$ mask (5.11). Third row output of the advanced filter with $f_{\Delta 3n}$ mask (5.13) and the fourth row present the output of the filter with f_{Lap} mask (5.18).

5.4.3 Complex Filter Structure

The past approaches based on the idea, that events are characterized by different values of the input signal. The differences can be calculated by the gradient approach. There are no mathematically specified demands were made for the behavior of the operators. In this section, a method is presented which is based on detailed mathematical formulations.

5.4.3.1 Approach of Canny

Canny (Canny, 1986) has formulated three assumptions for the development of his filter structure. These are

- Good detection, there should be a high probability to detect real events and a low probability to detect non-events.

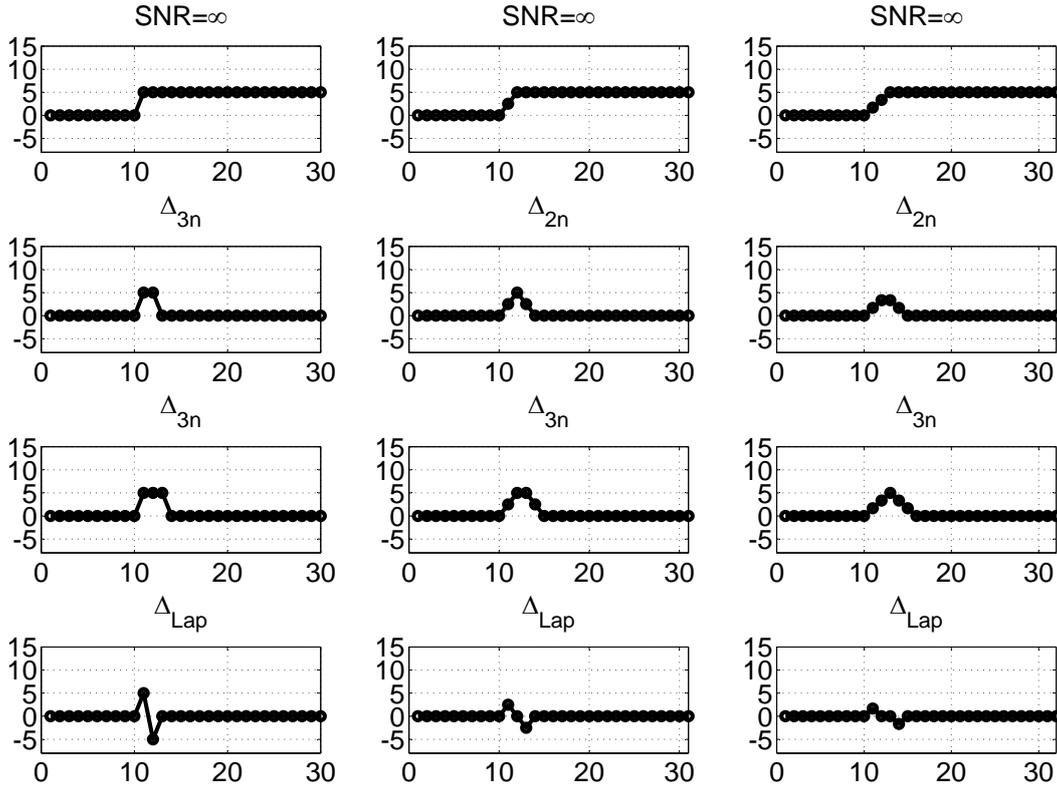


Figure 5.7 – Output of the filters for different gradients of the events.

- ▶ Good localization, the detected point should be as close as possible to the center of the real event.
- ▶ Only one response to a single event, this is close to the first criterion when there are two responses to the same event, one of them must be considered false.

The filter structure is dependent on the analyzed waveform. It is dependent on the signal form. So whether it is symmetric or antisymmetric. This can take some simplifications which lead to the equations given below. The previously described three sets of good detection, localization and error minimization are used to find the optimal filter. Canny defined the SNR, the localization and error minimization. The optimum filter is a correlation filter. The SNR is calculated as

$$SNR(t) = \frac{\left| \int_{-W}^{+W} p(-t)f(t) dt \right|}{n_0 \sqrt{\int_{-W}^{+W} f^2(t) dt}} \quad (5.21)$$

Where $f(t)$ is the impulse response of the filter. $p(t)$ is the input signal of the filter and describes the respective step. The window width is $2W$ and returns the length of the impulse response of the filter. In the denominator of the square value of the noise is calculated on the assumption that the impulse response of the filter is finite and that the noise is AWGN. This is then multiplied by an arbitrary noise coefficient n_0 . Using the convolution integral and the assumption that the stage is at $t = 0$, will be

calculated the response of the filter to the level in the counter. The good localization is calculated as

$$Localization(t) = \frac{\left| \int_{-W}^{+W} p'(-t)f'(t) dt \right|}{n_0 \sqrt{\int_{-W}^{+W} f'^2(t) dt}} \quad (5.22)$$

Eq. (5.22) calculates the distance between the real and the detected area. The goal is now to minimize this distance. For the derivation the functions which have already been used in the calculation of the SNR is used. To arrive at an optimal result, the product of the two equations 5.21 and 5.22 must be maximized:

$$\frac{\left| \int_{-W}^{+W} p(-t)f(t) dt \right|}{n_0 \sqrt{\int_{-W}^{+W} f^2(t) dt}} \cdot \frac{\left| \int_{-W}^{+W} p'(-t)f'(t) dt \right|}{n_0 \sqrt{\int_{-W}^{+W} f'^2(t) dt}} \stackrel{!}{=} max \quad (5.23)$$

Because the noise power density in the enumerator n_0 can not be minimized, the denominator must be maximized. This is satisfied if the filter function corresponding to the same input signal. This is fulfilled by the condition:

$$f(t) = G(-t) \text{ in } [-W, W] \quad (5.24)$$

The waveform of the filter was then determined empirically as a function of the input signal by Canny. He has examined different filter functions and comes to the conclusion that the use of the first derivative of the Gaussian function is the most suitable filter. However, there is the problem that any point which is higher than its neighbor, is detected as an event. This means that there are too many false detections due to noise. In this regard, the method has weaknesses.

5.4.3.2 Improvement of Canny Filter by Perona

An improvement of the Canny approach is proposed by Perona (Perona and Malik, 1990). He used a quadratic filter structure and not a simple one as Canny. In addition, two filters with the filter functions f_1 and f_2 are used, from which the output signals, are added. Each of these filters is tuned to a different part of the signal. A filter is responsible for the odd part. This reacts on stairs or rectangular functions. The other filter responds to the even part. These are, for example, peaks or ramp functions. The sum of the two outputs of the filters is squared. Formally, the output of the decision function is

$$\Delta P_{Perona}(t) = \sum_{i=1}^2 (p(t) * f_i(t))^2. \quad (5.25)$$

After several attempts, Perona came to the realization that for a even filter at best a twice derived Gaussian window $g(t)$ (5.26) is used (Perona and Malik, 1990).

$$\begin{aligned} g(t) &= \exp^{-\frac{(t-\mu)^2}{2\sigma^2}} \\ g'(t) &= -\frac{(t-\mu)}{\sigma^2} \exp^{-\frac{(t-\mu)^2}{2\sigma^2}} \\ f_1(t) = g''(t) &= \frac{1}{\sigma^2} \left(\frac{1}{\sigma^2}(t-\mu)^2 - 1 \right) \exp^{-\frac{(t-\mu)^2}{2\sigma^2}} \end{aligned} \quad (5.26)$$

This is a normal distribution with known parameters μ and σ . μ is the expected value of the function. With the predefined assumption, that the event is at point $t = 0$, the expected value becomes $\mu = 0$. Consequently, the only parameter of the filter is the standard deviation σ . This defines the width of the window. In a normal distribution it is assumed that by $\mu \pm 3\sigma$ 99,7% all values are mapped. For the convolution means that 99.7% of the energy of the window are included. For a practical application, the length of the window on is defined as $\pm 6\sigma$. The window length is therefore variable depending on the standard deviation being used. Perona founds out, that a filter structure with a standard deviation of $\sigma = [1 \ 3]$ has the best results. So in this thesis these three values are also for the mask of the filter. Fig. 5.8 shows the masks of the filter for different the standard deviations in the interval of $\sigma = 1, 2$ and 3. For the second odd filter f_2 Perona has transformed the first filter f_1 by the Hilbert

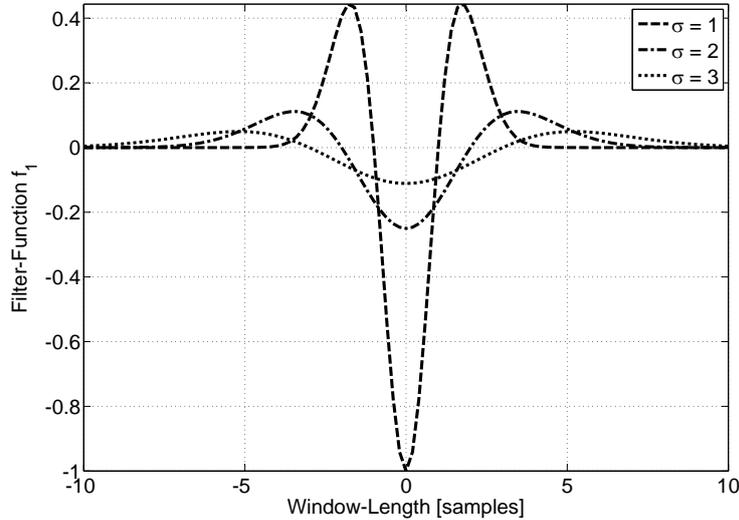


Figure 5.8 – Function of the first filter with different standard deviation σ . The mask is the second derivative from a Gaussian function.

Transform. The results were included in the basic structure of the approach of Canny (Canny, 1986).

$$f_2(t) = H \{f_1(t)\} = H \left\{ \frac{1}{\sigma^2} \left(\frac{1}{\sigma^2}(t-\mu)^2 - 1 \right) \exp^{-\frac{(t-\mu)^2}{2\sigma^2}} \right\} \quad (5.27)$$

The imaginary part of the Hilbert Transform of filter curve $f_2(t)$ is shown in Fig. 5.9.

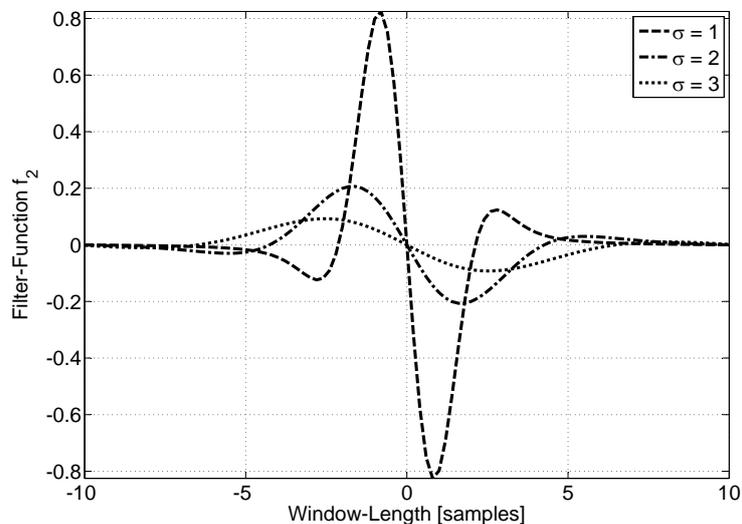


Figure 5.9 – Function of the second filter with different standard deviation σ . The mask is the Hilbert transform of the second derivative from a gaussian function.

Fig 5.10 shows the approach of Perona as a block diagram. The parameters of the individual components are shown. After the decision function, the result of the addition of ΔP_{Perona} is passed to a classification function. This then leads analogously to the first ED algorithm by the classification of the event.

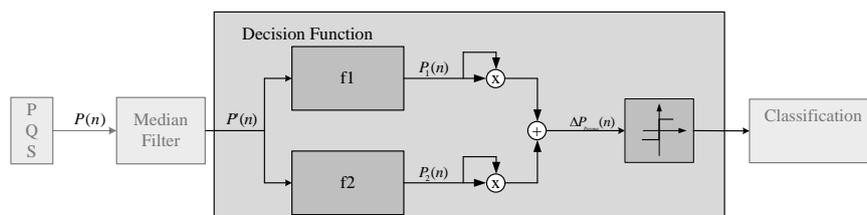


Figure 5.10 – Block diagramm of the approach from Perona with parameters.

Under the assumption that $\mu = 0$ the equations (5.26) and (5.27) becomes

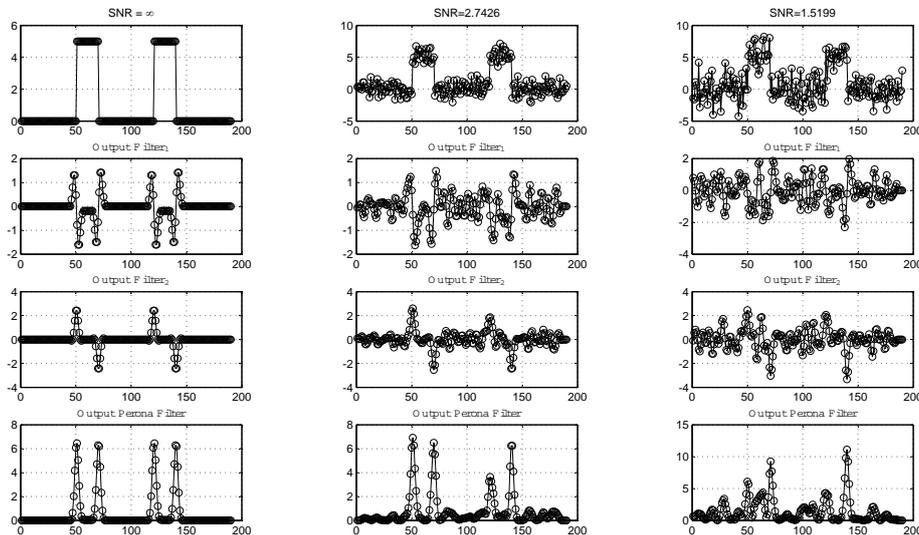
$$f_1(n) = \frac{1}{\sigma^2} \left(\frac{n^2}{\sigma^2} - 1 \right) \exp^{-\frac{n^2}{2\sigma^2}}, \quad (5.28)$$

$$f_2(n) = H \left\{ \frac{1}{\sigma^2} \left(\frac{n^2}{\sigma^2} - 1 \right) \exp^{-\frac{n^2}{2\sigma^2}} \right\}. \quad (5.29)$$

These equation were implemented and simulated.

5.4.3.3 Simulation of Peronas Filter

Analogous to the previously described high pass filter simulations the properties of the Perona filter were simulated. As noise AWGN is assumed again. This is superimposed to the useful signal with amplitude 5 W. Different standard deviations σ were used. Fig. 5.11 shows the simulation results. In the first row the signal with additive WGN are shown. The second row shows the output of the first filter with second derived Gaussian kernel. In the third row the output of the second filter is shown. Here the Hilbert transform of the first filter structure are the filter kernel. The last row shows the output of the decision function (cf. eq. (5.25)). The outputs of the filter show a good detection of the events. Also with decreasing SNR detects the filter events and suppresses the noise. For this simulation the center value for the standard deviation of the suggestion from Perona $\sigma = 2$ was used.



(a) An idealized waveform with $SNR \rightarrow \infty$. (b) The signal with AWGN with $SNR = 2,74$. (c) The signal with AWGN with $SNR = 1,51$.

Figure 5.11 – Simulation of the Perona Filter. The standard deviation of the filter is $\sigma = 2$. The first row presents the signal with AWGN. In the second and third row the outputs of the two filters are shown (cf. (5.26), (5.27)). The last row shows the output of the decision function (cf. (5.25)).

For a distinction whether it is a switching on or switching off event, the filter is limited. By squaring the sum of the two outputs (cf. eq. (5.25)) this information is lost. The output values of the filter are positive in both cases. For a practical solution, additional information must be used. It would be possible to use the output of one of the two individual filters f_1 or f_2 . These have different signs for the outputs of the two events.

In Fig. 5.12 the behavior of the output of the filter is shown for different gradients

of the switching event. Again, the three gradients $\Delta = \frac{5}{1}, \frac{5}{2}, \frac{5}{3}$, which have 1,2 and 3 samples were simulated analog to the first simulation. It can be seen that the filter detects the events well. However, the amplitude of the output of the filter decreases with a decreasing gradient.

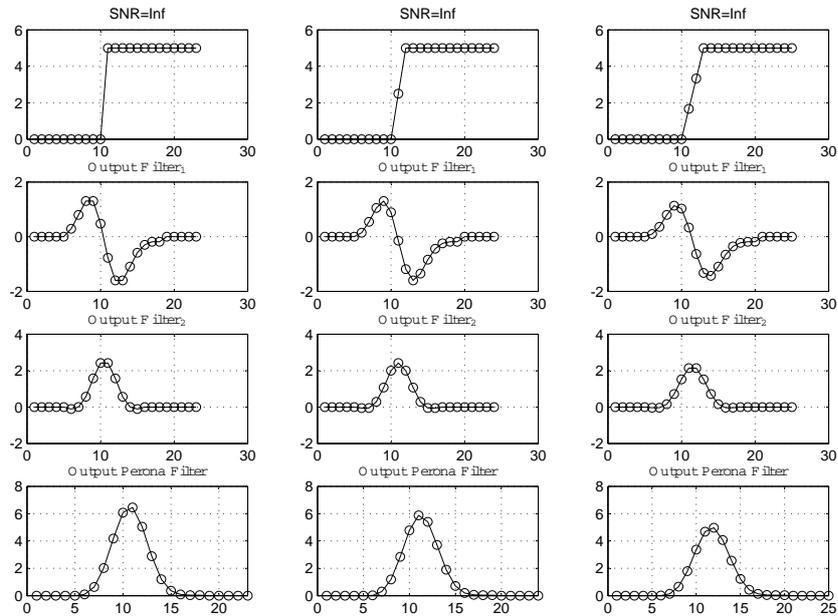
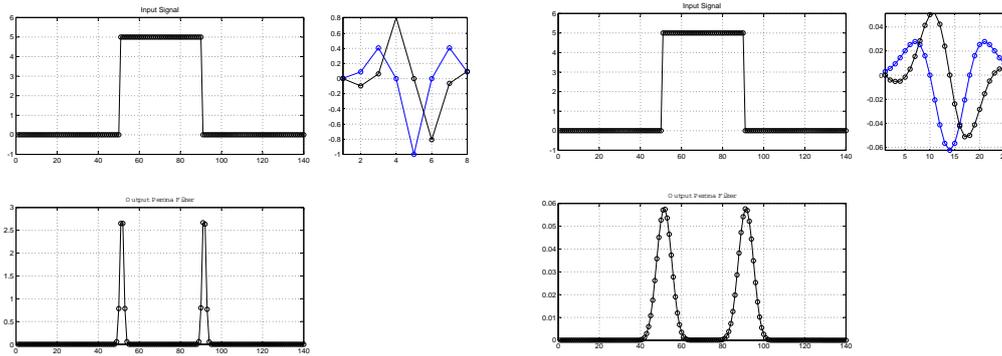


Figure 5.12 – Output of the filters for different gradients of the events.

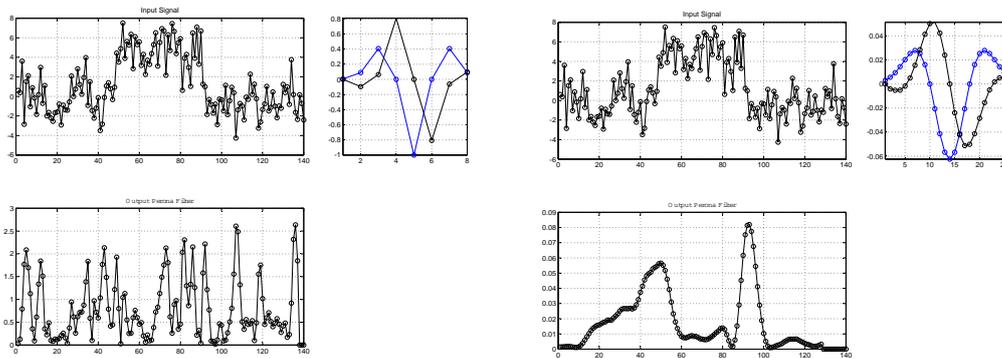
Canny and Perona have taken during their derivation two essential assumptions. A good detection and the best possible noise reduction. However, an optimization of both to its maximum is not possible. Either one has a good detection of the events or a good noise reduction. In practice, therefore, an intermediate value must be found. In the following the dependence of detection and noise reduction has been simulated.

Fig. 5.13 shows an ideal signal without noise. Two filter functions were simulated based on the findings of Perona. For one, it is a narrow filter with $\sigma = 1$. This filter is shown in Fig. 5.13(a). It is shown the input signal and the output of the filter. On the right side the mask of the two filters are shown. This filter structure is very narrow and detects the events very precisely. The same simulation but now with a standard deviation of the filter mask of $\sigma = 3$ is shown in Fig. 5.13(b). This filter function is much broader. As can be seen, the output function of the filter is broader. Thus, the simulation meets consistent with the assumption and derivation. In Fig. 5.14 the behavior of the filter for a noisy signal is shown. There the two filter masks with $\sigma = 1$ and $\sigma = 3$ was used again. Now, it can be seen that the narrowband filter has a lower noise reduction with decreasing SNR. Figure 5.14(a) clearly shows that the two events lost in the noise. This contrasts with the broad-band filters. The two events are still detectable.



(a) Output of the Perona filter for a filter function with $\sigma = 1$. (b) Output of the Perona filter for a filter function with $\sigma = 3$.

Figure 5.13 – Simulation of the Perona Filter with different length of the filter, depended from the standard deviation σ of the gaussian function. Here an idealized signal was used, were the $SNR \rightarrow \infty$.



(a) Output of the Perona filter for a filter function with $\sigma = 1$. (b) Output of the Perona filter for a filter function with $\sigma = 3$.

Figure 5.14 – Simulation of the Perona Filter with different length of the filter, depended from the standard deviation σ of the Gaussian function. Here an idealized signal was used, were the $SNR = 1,43$.

5.4.4 Short Time Fourier Transformation

The approaches described before are use the time series signal for the event detection. Another possibility to detect events in a non stationary signal is a time frequency analyzing method. For this purpose, the time signal must be transformed into the frequency domain.

In chapter 2.4 some different methods were described to analyze a signal with whose spectrum. By deriving the high-pass filter it has been described that abrupt changes in the signal have high-frequency components. It should therefore be examined whether these high-frequency spectral components can be used for event detection.

A possibility of transformation the signal in frequency domain provides the FT. But

the time series is a non-stationary signal. While the FT does not provide information about the temporal occurrence of frequency components in the signal, it can't be used for the transformation. The STFT is suitable for non-stationary signals where the frequency characteristics change over time. Instead of calculating the spectrum of the entire time signal, shorter time segments are converted into the spectral range. For this, the signal $p(t)$ is multiplied with a window function w and calculates the spectrum for that signal.

$$STFT(\tau, f) = \int_{-\infty}^{\infty} p(t)w^*(\tau - t)e^{-j2\pi ft} dt \quad (5.30)$$

τ is the observed moment of the signal. In the most cases a real window w can be used instead of the conjugate complex w^* . The STFT transforms the time series into frequency domain and presents the spectral components over time. The time-discrete form of the STFT is

$$STFT(m, f) = \sum_{n=-\infty}^{\infty} p(n)w(n - m)e^{-j2\pi fn}. \quad (5.31)$$

After the calculation of the spectrum of the signal at time τ , the Fourier coefficients are added.

$$\Delta P_{STFT}(\tau) = \sum_{f=1}^N STFT(\tau, f) \quad (5.32)$$

N is the number of coefficients and depends of the window length.

5.4.4.1 Simulation of the STFT

The first simulation was done with an idealized signal. As shown by the simulations of the two filters previously simulated the SNR goes to $SNR \rightarrow \infty$. It will examine the extent to which the parameters of the STFT impact on the detection of events. As in chapter 2.4.2 described, the resolution in the time domain and the resolution in the frequency domain are disproportionately. This depends on the length of the window ω .

As window function a Hamming window was used. This reduces the leakage-effekt. Fig 5.15 shows the results of the simulation for a Hamming window of size 10. In the first row on the left side, the input signal is shown. An idealized signal without noise and an amplitude of 5 W. 4 switching events (2 on and 2 off) are shown. In the second row the results of the STFT is shown. It is shown the spectrogram. Here, the power density spectrum (cf. chapter 2.4.2.2) over time and the frequency is plotted.

It is well shown, that the intensity of the spectral components increase when a appliance is switched on. Based on eq. (5.32) the delta value is calculated. Because only the

upper frequency components contain information about abrupt changes in the signals, only the high-frequency spectral components were added. The result is presented in the last row of Fig. 5.15. By choosing the length of the narrow window 10, the detection of the events in the time domain is relatively well. If the width of the window increases

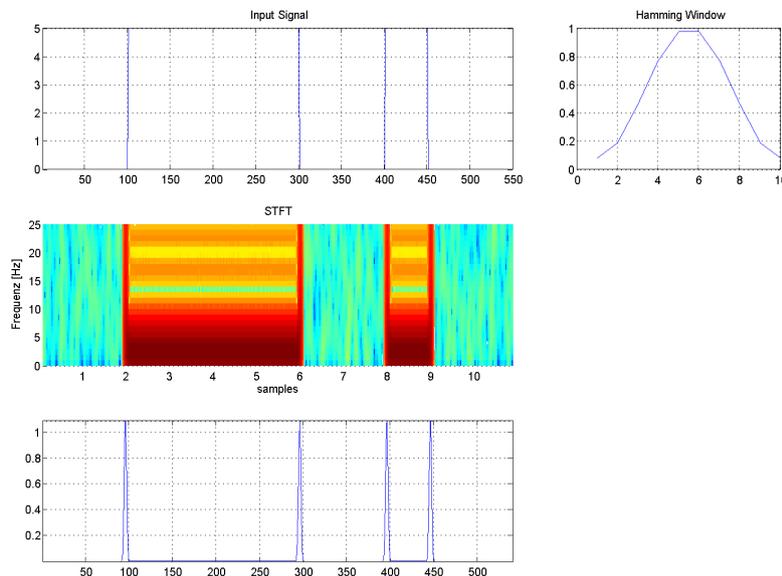


Figure 5.15 – Simulation result for the STFT with a Hamming window of size 10. The input signal is idealized without noise.

the detection would be decreasing. Therefore a better spectral resolution is achieved. In Fig. 5.16 the results of an expanded window size with length 40 are presented. Now, it can be seen that the time resolution is poor. However, in the spectral representation of the STFT, the energy is better resolved. Even in the simulation of the Perona approach 5.4.3.3 was found that a compromise between resolution in the time domain and resolution in the frequency range must be taken here.

It is also noted that this approach does not distinguish between one and off events. This property which is of importance in the time domain does not occur in the frequency domain. Again for the distinction between on and off events secondary information must be used. One possibility would be to investigate the spectral value of the constant component of the detected event. Is this higher than before, it is an on event. Is this less an off event has been detected.

In the second simulation, the noise component was increased. The standard deviation of the WGN was chosen so that $SNR = 3dB$.

In Fig. 5.17 the simulation result of the STFT with a Hamming window of size 10 is shown. It can be seen that the noise now has significant influence on the detectability of events. It can be seen that with a suitable threshold the events are detected. However, the disturbances caused by the noise are increasing.

The simulation results of the STFT with a wider window for a noisy signal are shown in Figure 5.18. This shows the advantage of the wider window compared to the narrower.

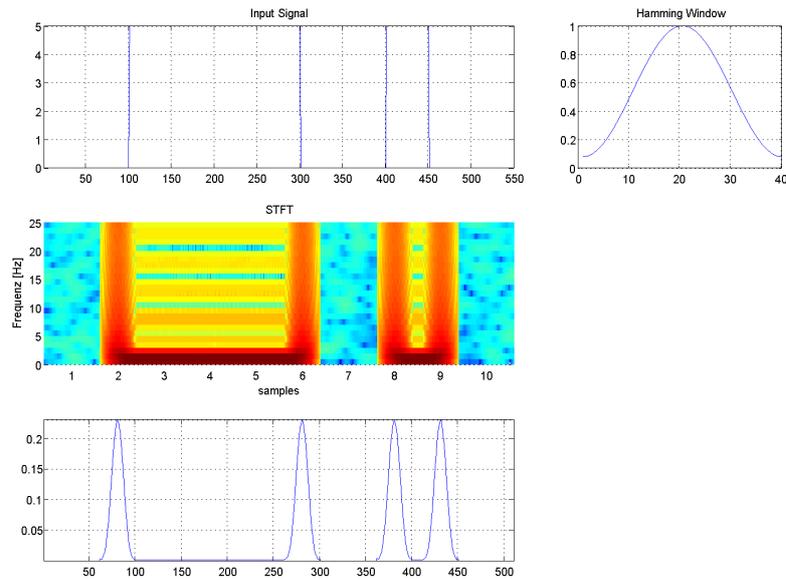


Figure 5.16 – Simulation result for the STFT with a Hamming window of size 40. The input signal is idealized without noise.

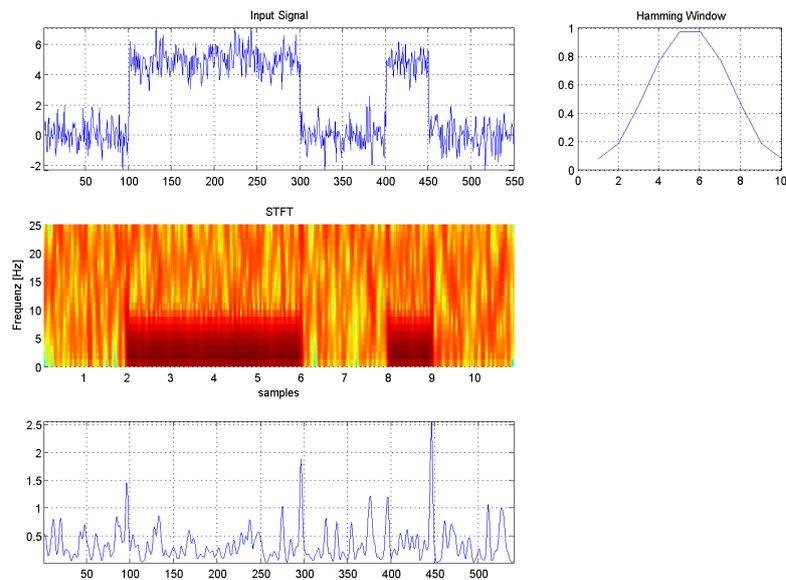


Figure 5.17 – Simulation result for the STFT with a Hamming window of size 10. The signal was superimposed by noise. It is shown, that the events can be detected. But the influence of noise is increasing and FP are detected.

Although the localization of the events through the wide window is difficult but this has a better noise reduction. The events are detected and the noise are reduced.

In the previous two simulations (HP and Perona filter) was still under investigation, the behavior of the filter at a smaller gradient of the on events. It is clear that a smaller gradient has less high-frequency components. Consequently, the energy, which

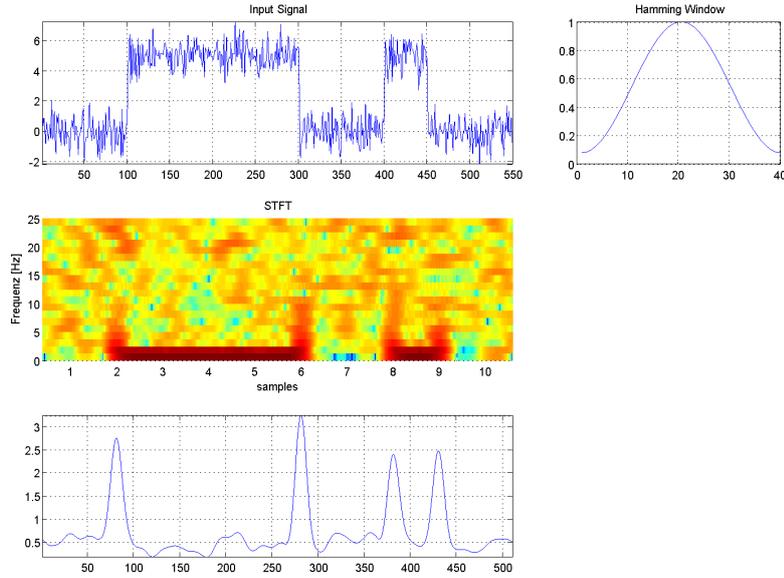


Figure 5.18 – Simulation result for the STFT with a Hamming window of size 40. The signal was superimposed by noise. It is shown, that the events are detected.

is in the higher harmonics decrease. The simulation results of the STFT for different gradients of the switching on event are shown in Fig. 5.19. As can be seen reaches the energy by a gradient of $\Delta = \frac{5}{1}$ a value of 1.2. This is shown in the first column. If the gradient is halved ($\Delta = \frac{5}{2}$) the energy is decreasing by 2/3. As shown in the second column. In the last column the gradient is $\Delta = \frac{5}{3}$. Now the energy has approximately 10 % of the energy of the $\Delta = \frac{5}{1}$ gradient. This is a major disadvantage of this method. If the threshold is adapted to smaller gradients, the effect of the noise for the detection will be increasing. This will lead to false detections.

The simulations show that this approach will be used under certain conditions, to detect events. The choice of the window length is a crucial optimization criterion. This will affect the detect ability and the suppression of the noise. Depending on the threshold, steep gradients can therefore be readily detectable. Wherein decreasing gradients are quickly lead to false detections.

5.4.5 Classification Function

In the past sections three algorithms for the decision function of the event detector were presented. The second block of the event detector includes the classification function (cf. Fig. 5.4). The decision function output values of the three approaches are defined as:

- ▶ $\Delta P_{HP}(t)$, as output of the High-Pass Filter (generalized for the 3 masks $f_{\Delta_{2n}}, f_{\Delta_{3n}}$ and f_{Lap})
- ▶ $\Delta P_{Perona}(t)$, as output of the Peronas Filter

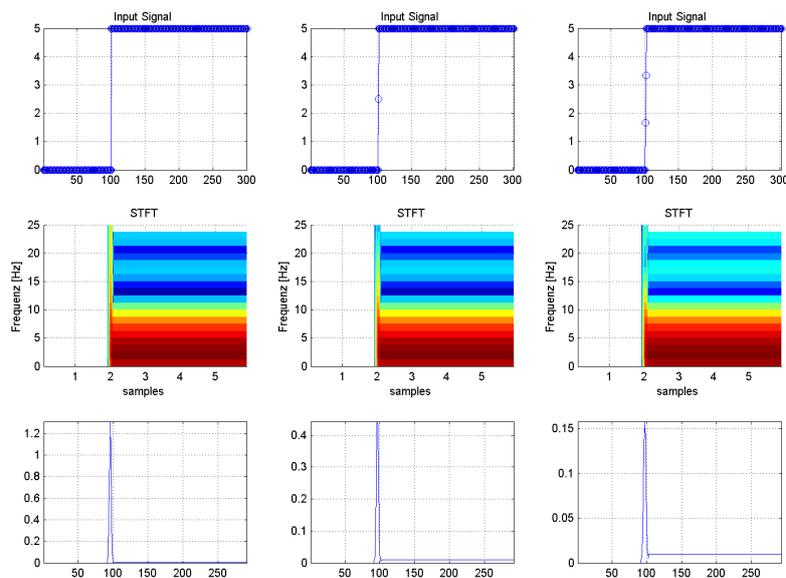


Figure 5.19 – Simulation results of the STFT for different gradients of the switching on event. On the left side the results for a gradient with $\Delta = \frac{5}{1}$ is shown. In the middle the gradient is $\Delta = \frac{5}{2}$ and on the right side the gradient is $\Delta = \frac{5}{3}$.

- $\Delta P_{STFT}(t)$, as output value of the Short-Time Fourier Transformation STFT

These values are used as input for the classification function to label an event or no event.

5.4.5.1 Threshold for the Classifier

The simulations have shown that the output values of the decision functions have different ranges. Now we need to interpret these values in this case to classify them. In chapter 2 a simple classifier was presented. It is a threshold decision maker.

If $\Delta P(t)$ (generalization for $\Delta P_{HP}(t)$, $\Delta P_{Perona}(t)$ and $\Delta P_{STFT}(t)$) is greater or smaller than a defined threshold ΔP_{th} a switching on or off event at time t is detected.

$$SwONev(t) = \begin{cases} 1, & \Delta P(t) \geq \Delta P_{th} \\ 0, & \text{else} \end{cases} \quad (5.33)$$

$$SwOFFev(t) = \begin{cases} 1, & \Delta P(t) \leq -\Delta P_{th} \\ 0, & \text{else} \end{cases} \quad (5.34)$$

It should be noted that only the approach using a high-pass filter can distinguish between falling and rising events. Eq. (5.34) presents the general case.

The definition of the threshold plays an important role for the quality of the event detector. If it is too large, there is a risk that events are no longer detected. The TPR decreases. If the threshold is set too small, the TPR are increase. But faults which are caused by noise also detected as events. Consequently, the False Positive Rate (FPR) also increases and so the Positive Predicted Value (PPV) decreases. This dilemma was described in chapter 2.3. The choice of ΔP_{th} thus determines whether it is a liberal or conservative classifier.

5.4.5.2 Simulation of the Classifier

In this section the dependence of the detection rate from the threshold is simulated. As decision function an high pass filter with $f_{\Delta_{3n}}$ mask is used. For the simulation different amplitude values were defined for the input signal. The amplitudes are $A_1 = 5$ W, $A_2 = 10$ W and $A_3 = 15$ W. The ideal signal was superimposed by WGN. In the first row of Fig. 5.20 the input signal is shown. The standard deviation of the noise is $\sigma = 1.778$. For different amplitude values different SNR values are calculated (cf. eq. (4.17)). These are $SNR_{A1} = 1,48$, $SNR_{A2} = 4,49$ and $SNR_{A3} = 6,25$. The output signal of the decision function is shown in the second row in Fig. 5.20. It is shown, that the noise amplitudes have values over 5 W.

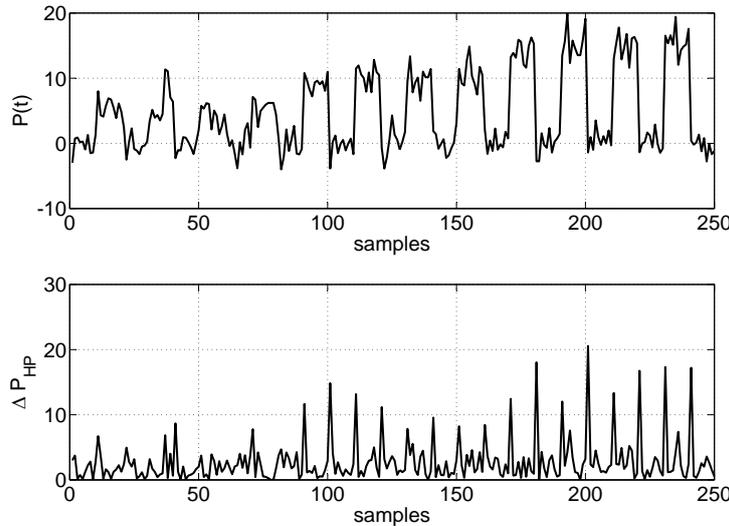


Figure 5.20 – Input and output signal of the decision function. For the mask of the high pass filter the $f_{\Delta_{3n}}$ mask is used. The input signal has 3 amplitude values $A_1 = 5$ W, $A_2 = 10$ W and $A_3 = 15$ W. With a standard deviation of $\sigma = 1.778$ the SNRs for the different amplitudes are $SNR_{A1} = 1,48$, $SNR_{A2} = 4,49$ and $SNR_{A3} = 6,25$.

The output signal was fed to the threshold decision makers. With increasing threshold value ΔP_{th} the TPR is decreasing. This is shown in the first row in Fig. 5.21. The two threshold crossings at 5 W and 10 W can also be seen. If all events should be detected, the threshold must be minimized. In the second row of Fig. 5.21 correlation between the PPV and the TPR is shown. By increasing the TPR the PPV is decreasing. For

a practical application a threshold must be found where the TPR is still high but the influence of noise is not yet large. This can maybe the break even point (cf. chapter 2.3.2.3).

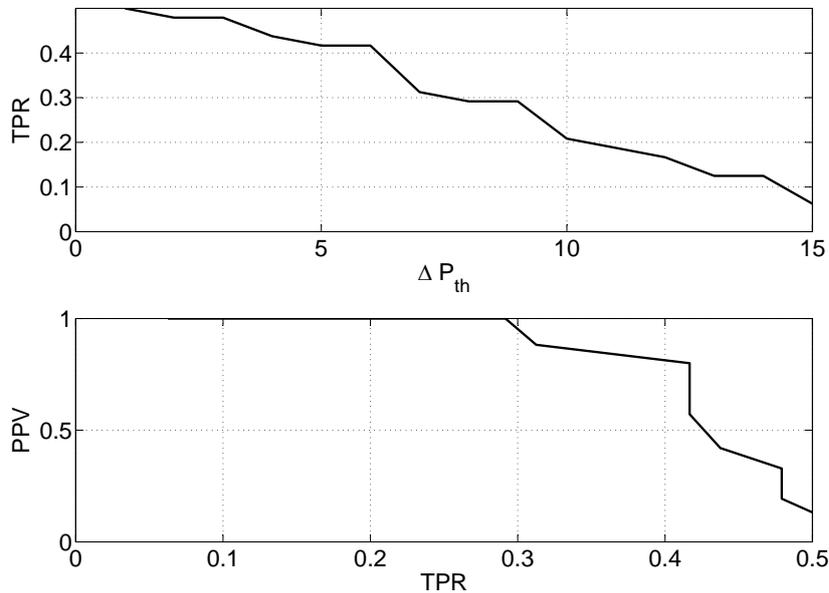


Figure 5.21 – Simulation results of the classifier function. In the first row the TPR in dependence from the threshold ΔP_{th} is shown. In the second row the PPV over the TPR is plotted.

In this section three approaches for the decision function of the event detector are described and simulated. For the classification function the threshold decision makers was described. In the following section the approaches are analyzed by real measurements.

5.5 Results for the Event Detection

This section contains the statistical evaluation of the quality of the event detector. For this the approaches are tested by real measurements. The individual process will be compared with each other. At the end of the section an approach is defined which can be implemented in a later smart meter system.

5.5.1 Experimental Procedure

For the evaluation of the algorithms the measurements described in chapter 4.5.2.2 were used. These are measurements of 10 different appliances. These have a power consumption between 3.8 VA and 1022 VA. Overall 10,000 switching cycles were recorded. The switching times are subject to a uniform distribution. Each appliance is thus switched on and off by 500 times.

All previously described methods have been implemented. The methods were tested with different parameters. Different thresholds ΔP_{th} has been defined for the classification function. The analysis was repeated several times. To compare the results with a known method of NALM the approach of Hart was also analyzed. With the resulting the following statements can be made

- ▶ Analyze the PPV at a constant TPR
- ▶ comparing the absolute values of each method at the break even point
- ▶ analyze the FPR at the maximum TPR

5.5.2 Best Signal for the Event Detection

The signals which can be calculated from the measured signals $u(t)$ and $i(t)$ are the active power (P), the reactive power (Q) and the apparent power (S) (cf. 4.2.1). These signals are calculated and pre-filtered in the block measurement system. These are the input values of the event detector. Now the best signal in sense of detection and noise cancellation should be analyzed.

An example of a real measurement is shown in Fig. 5.22. In the first row the power signals (P, Q and S) are presented. Where the first one is Q, the second is P and the last one is S. In the second row the results of the filter output are shown. For this example the HP filter with $f_{\Delta_{3n}}$ mask is used.

The evaluation of the results has shown that the reactive power as input signal provides the worst results. The reason is that some of the measured appliances have no or negligible small reactance. Therefore, some switching on or off events have no reactive power. But every power signal is superimposed by noise. The reactive power has the worst SNR. From this reason the reactive power can't used for the event detection as input signal. The other signals P and S have similar results. So both can be used as input signals. For the further work the active power P is used.

5.5.3 Results of the Event Detection Methods

In that section the three approaches are analyzed with real measurements. Fig. 5.23 shows an example of a real measurement and the outputs of the different algorithms. In the first row the power signal P is shown. Different appliances were switched on and off. In the second row the results of the three HP Filters are shown. In the third row the output of the filter with Perona mask are presented. And in the last row the sum off the higher harmonics calculated with the STFT are shown. The analyze were done for all switching events.

A first method to compare the different event detection algorithms is using the break even point. At this point the values of the TPR and PPV can be analyzed. Fig. 5.24 shows the PPV over the TPR of the three approaches. The threshold of all three approaches ΔP_{th} were decreased. So with a smaller ΔP_{th} the TPR increase but analogously the PPV are decrease. This is because more and more disturbances are

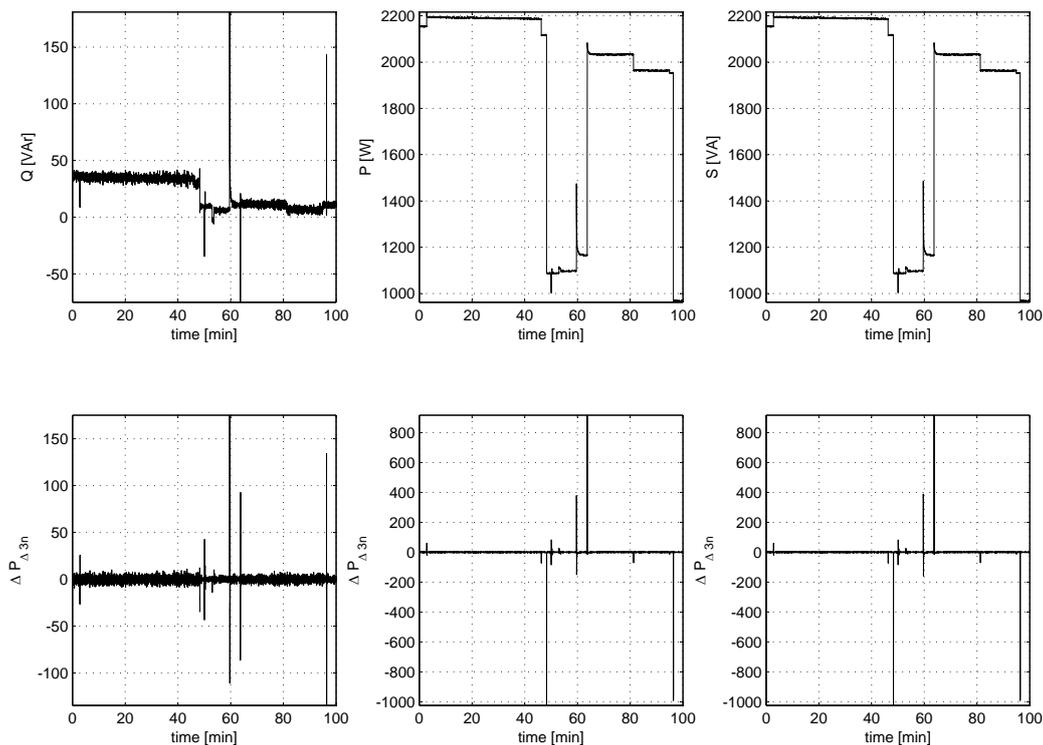


Figure 5.22 – Example of a real measurement. In the first row the power signal Q, P and S are shown. In the second row the output of the event detector using a high pass filter with $f_{\Delta_{3n}}$ mask for the three different input signals is shown.

detected as an event. The better a method, the slower fall the PPV with increasing TPR.

Fig. 5.24(a) shows the results of the HP Filter-Approach. Three different masks are used, a Laplace-Operator $[1 \ -2 \ 1]$, a $f_{\Delta_{2n}}$ mask $[1 \ 0 \ -1]$ and a $f_{\Delta_{3n}}$ mask $[1 \ 0 \ 0 \ -1]$. The best results are reached with the $f_{\Delta_{2n}}$ - and $f_{\Delta_{3n}}$ -mask. Because they have the biggest area under the curve. Also the break even point is much higher as for the Laplace mask. Fig. 5.24(b) shows the results of the Perona-Approach. For the derivate Gaussian windows, different σ values are used. For our signals the best value are $\sigma = 2$. The last Fig. 5.24(c) shows the results for the STFT. A Hamming-window with different length w was used. Is the window to small, the noise has a high influence to the PPV. Is the window to large, the noise is reduced, but not all events are detected. The best window length for our signal is $w = 50$.

The optimal parameters for the measurements are used to compare the three methods now. The three approaches are verified with these parameters. The results are shown in Table 5.1. The table shows the different detection rates for the approaches; Hart, HP-Filters ($f_{\Delta_{2n}}$ -mask, $f_{\Delta_{3n}}$ -mask and f_{Lap} -mask), Perona-Filter and the STFT. For different TPR values the PPV values are presented. From these relativ rates the absolute True Positives (TP) and FP values are calculated (cf. eq. (2.2) and (2.4)).

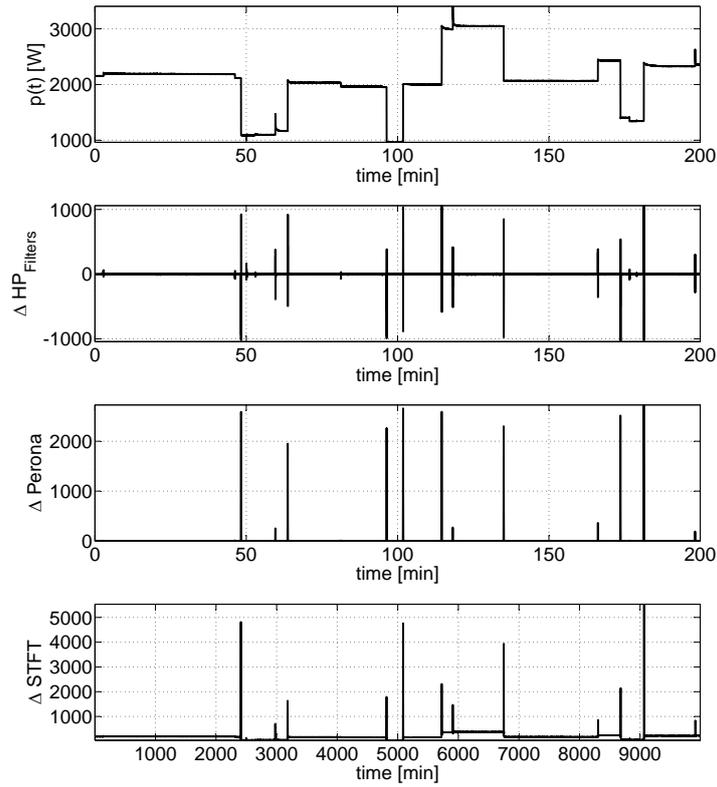


Figure 5.23 – Outputs of the three approaches for a real measurement signal. In the first row the signal P is shown. In the second row the outputs of the HP Filters, in the third row the output of the Perona Filter and in the last row the output value of the STFT approach is shown.

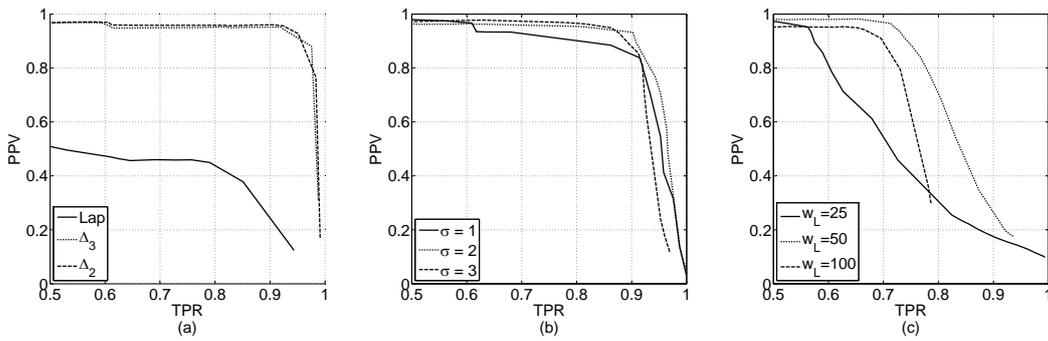


Figure 5.24 – PPV over TPR for the three approaches to find the optimal parameter values. (a) Filter-Approach with three different masks, (b) Perona-Approach with three different σ values and (c) STFT-Approach with three different window length.

These values are shown in the last columns of the table.

The usage of an event detector depends on the application. The best results show the Perona-Filter. With this filter it is possible to detect all events. By decreasing

Approach	Mask/ Parameter	TPR	PPV	N_e	TP	FP
Hart	(1 -1)	85 %	90 %	10000	8500	944
		90 %	82 %	10000	9000	1975
		95 %	70 %	10000	9500	3875
		100 %	5 %	10000	10000	180500
$f_{\Delta_{2n}}$	(1 0 -1)	85 %	92 %	10000	8500	944
		90 %	85 %	10000	9000	1588
		95 %	75 %	10000	9500	3000
		100 %	8 %	10000	10000	109250
$f_{\Delta_{3n}}$	(1 0 0 -1)	85 %	92 %	10000	8500	739
		90 %	90 %	10000	9000	1000
		95 %	80 %	10000	9500	2250
		100 %	10 %	10000	10000	85500
f_{Lap}	(1 -2 1)	85 %	87 %	10000	8500	1270
		90 %	70 %	10000	9000	3857
		95 %	45 %	10000	9500	11000
		100 %	3 %	10000	10000	307166
Perona	$\sigma = 2$	85 %	95 %	10000	8500	447
		90 %	94 %	10000	9000	574
		95 %	72 %	10000	9500	3500
		100 %	20 %	10000	10000	38000
STFT	$\omega = 50$	85 %	38 %	10000	8500	13868
		90 %	28 %	10000	9000	23142
		95 %	10 %	10000	9500	81000
		100 %	0.1 %	10000	10000	9080910

Table 5.1 – Results for the event detection methods.

the threshold more and more TP are detected. This approach has also the best noise cancellation. By detecting 100 % of all TP the PPV has a value of 20 %. The second best method is using a high-pass filter with $f_{\Delta_{3n}}$ mask. This has the advantage that it needs less space because of the number of values of the mask. Together with the high pass filter with $f_{\Delta_{2n}}$ mask these approaches are better as the events detection method presented by Hart. So it was shown that it is possible to improve the quality of the event detector. More events were detected by reducing the influence of noise. The high pass filter with f_{Lap} mask and the approach using the STFT are not suitable for our problem. These methods have a poor noise cancellation and cannot be used in a smart meter system.

5.6 Conclusion

In this chapter the event detection as a preliminary calculation for the classification was described. An event detector was developed. Based on the work of Hart an improved event detector was developed. This detector should have a higher detection rate and a greater noise cancellation as the approach of Hart. This detector consists of two blocks.

These are a decision function and a classification function. For the development of the decision function algorithms different assumption were defined. The event detector should detects events of appliances with a power consumption between 5 W-3680 W. Abrupt changes in the times series with gradients of 5 W should be detected. The last assumption defines the decreasing faults which results from noise.

The signals passed to the event detector were analyzed. From the measurement we know that the signal is superimposed by noise. Based on these disturbances an investigation of the state of art of event detection methods was done. Different methods from NALM, image processing and other fields of signal processing were described.

The main part of that chapter was the development of event detection algorithms. Three approaches were analyzed in detail.

The first one using a high-pass filter structure. Different masks were analyzed which based on the gradient and the Laplacian operator. The behavior of the three mask $f_{\Delta_{2n}}$, $f_{\Delta_{3n}}$ and f_{Lap} were simulated. For this an idealized signal were used. This signal was superimposed by noise. Also different gradients of the switching on event were simulated. It was shown that the $f_{\Delta_{3n}}$ mask has the best results.

The second approach based on a more complex filter structure. It is an approach based on the work from Perona. A twice filter structure was used with a twice derivative Gaussian function as mask for the first filter and the Hilbert transform of this as mask for the second filter. This method was also simulated analogously to the simulation of the HP filters. It was also shown that it is not possible to optimize the detection and the noise cancellation.

The last method is different from the first and second approach. It doesn't use a filter structure. It based on a time frequency analyzing method. This approach using the STFT for an analyze of the harmonics of the switching on event. Because the abrupt events have higher components.

After the simulation of the approaches there were analyzed with real measurements. To have a good statistical statement 10.000 switching cycles of different appliances were used. It was shown, that the approach of Perona has the best results. It follows the filter mask derived from the gradient operator. Both approaches have better results as the approach of Hart. With the Perona mask the PPV is 12 % better as the approach of Hart for a TPR of 90 %. The detecting of TP is greater as the rates from Hart. Also these approaches have a better noise cancellation as the approach of Hart.

For a later implementation in a smart meter system one of the two filters with the mask of Perona or $f_{\Delta_{3n}}$ mask can be used.

6 Classification

Contents

6.1	General Description of Neural Networks	98
6.1.1	Biological Systems	98
6.1.2	Historical Overview	99
6.1.3	Applications for Artificial Neural Networks	100
6.2	Artificial Neural Networks	101
6.2.1	Perceptron	101
6.2.2	Transfer Function	102
6.2.3	Structure of Networks	102
6.2.4	Learning Algorithms	104
6.2.4.1	Learning Rule for the Perceptron	105
6.2.4.2	Learning Rule for the MLP	105
6.2.5	Design of a MLP	106
6.2.6	ADALINE	107
6.2.6.1	Structure of an ADALINE	107
6.2.6.2	Learning Rule ADALINE	108
6.3	ADALINE for the Classification	109
6.3.1	ADALINE for Parameter Estimation	109
6.3.2	Simulation	110
6.3.3	Results	112
6.4	MLP for the Classification	114
6.4.1	Features for the Classification	115
6.4.1.1	Train the MLP	116
6.4.2	Results of the Classification Method	116
6.5	Conclusion	118

Chapter 4 describes the measurement system. With that system the measurements for the development of the disaggregation algorithms were recorded. In the previous chapter 5, the event detection was described. This is used to detect the events of the appliances. In this chapter, the last block of disaggregation chain is described. This is the classification of appliances in the load profile.

If an event is detected, a classification is performed at this point. The classifier then assigns each event to a class of appliances. Fig. 6.1 shows the classification as part of the disaggregation chain.

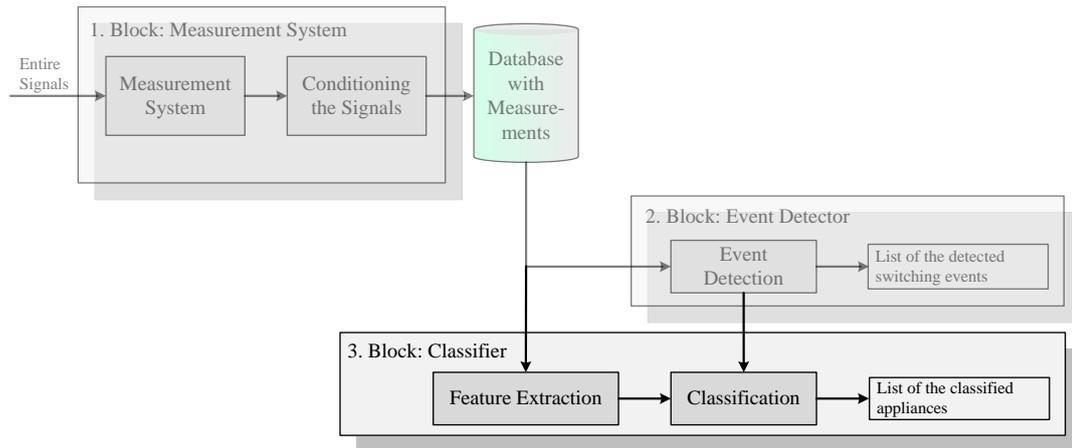


Figure 6.1 – Structure for the disaggregation of appliances in the load profile of residential buildings. Here the third block is described. The classifier assigns each event to a class of appliances.

From the recorded measurements the input vector and the output vector are known. For this reason, a supervised learning method is used. As classifier an artificial neural network is used. At the beginning of the chapter, a theoretical introduction is given in the area of ANN. For the classification different features can be used. In this work for the features the power signals P , Q and S plus derived signals are used. To extract features from the signal a special form of an ANN will be presented. At the end of the chapter, two methods are presented for classification. The one is used to classify a specific class of appliances. The structure of the ANN is simple. A more complex ANN is used to classify the appliances from the measurements (cf. chapter 4.5.2.2). For this purpose, different features are examined.

6.1 General Description of Neural Networks

6.1.1 Biological Systems

The human has a variety of cognitive abilities. These allow him to recognize complex issues and develop solutions. Much of the research in the field of ANN was inspired and influences by the knowledge of biological nervous systems. The fundamental part in biological systems is the so-called neuron. A neuron is a small cell which receives

electrochemical stimuli from several sources. To these inputs, it reacts with electric pulses. These are delivered in turn to other neurons. The human brain consists of about $10^{10} - 10^{12}$ neurons. These are connected in themselves. This high density makes the brain into a complex fast-responding processing unit. Neurons are composed of a cell nucleus, a cell body and several dendrites. These are connected via so-called synapses with the outputs of other neurons. The output of the neuron, the so-called axon, is also connected to other neurons. The schematic representation of a biological neuron is shown in Fig. 6.2.

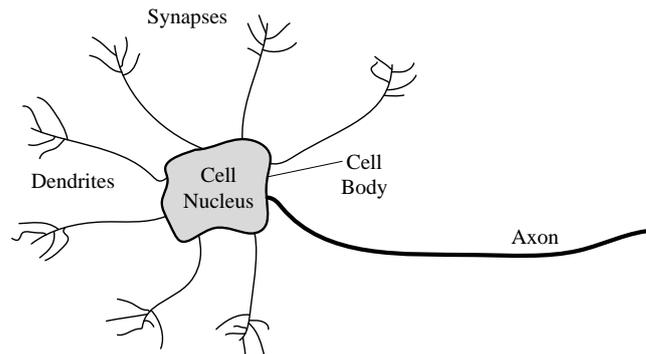


Figure 6.2 – Model of a biological neuron.

Neurons can be connected to thousands of other neurons. Exceeds the cumulative sum of the input signals a certain threshold, the neuron fires. This means it will send via the axon a series of action potentials. This impulse conduction can be executed in a few milliseconds. The learning of the cell activities is called metabolic growth. It affects the potential loads which generates a synapse. This can be thought of as a weighting. What will we find in analogy with the ANN.

6.1.2 Historical Overview

The basic structure of artificial neural networks is similar to nerve cells, so called neurons. The first approaches were founded by McCulloch and Pitts in 1943. They described a net which can calculate all logical and arithmetical functions. 1947, they point out that such a network can be used for spatial pattern recognition. Donald O. Hebb in 1949 formulated his Hebb learning rule, which represents almost all neural learning method in its general form. A famous work is from Rosenblatt (Rosenblatt, 1958). He developed the first neuron computer in 1958. The first adaptive neuron, which calculates the weights under a learning rule was developed by Widrow and Hoff (Widrow and Hoff, 1960) in the year 1960. This approach is described more in detail in section 6.2.6. This network reached as the first wide commercial distribution. Application found it in analog phones for real-time echo filtering. The neural network learned with the delta rule. 1969 Marvin Minsky and Seymour Papert gave a detailed mathematical analysis of the perceptron. They showed that important problems cannot be solved. Among other things, XOR operators cannot be resolved and there are problems in the linear separability. The result was a temporary end to the research in the field of neural networks. A renaissance reached the ANN in the 1980s. 1985

Hopfield published a solution of the traveling salesman problem by a Hopfield network. In 1985, the backpropagation learning method was developed as a generalization of the delta rule (Hopfield and Tank, 1985). Thus non-linearly separable problems through multi-layer perceptrons are solvable. Minsky's assessment was therefore refuted.

6.1.3 Applications for Artificial Neural Networks

ANN can be used for various applications. Here is an overview of the most common applications are given. The descriptions are general in nature and thus generic. The methods which are used in this thesis will be explained in more detail later in the chapter.

Some ANN structure can learn to take patterns and save these. If you then present them an associated pattern the stored one is recovered. Here, the associated pattern can be the same, can have slight differences or can have massively differences from the stored one. This kind of network is called **associative memory**.

ANN can be used for **data compression**. This is of interest when a large amount of data is to be transmitted or stored. Here compute the ANN a mapping. This makes a reduction of the input pattern. Through this compression pattern of the n -dimensional space are transformed in the m -dimensional space, where $m < n$.

ANN can be used to assign an input pattern to a corresponding output. For example, the different symptoms can then connect to an illness. The **diagnosis** of complex systems is a common application area for expert system.

A **prediction** is used in many areas. An application example is the prediction of electricity consumption. Thus, the power company can plan the energy reserves better. The output of the ANN can have three states. First, something happens or not; second, when something happens; or third, the probability of something happening. For an acceptable accuracy a sufficient amount of examples is required for the ANN.

When data are combined from different sources, it is called **multisensory data fusion**. This process includes determining, associating, correlating, estimating and combining of data. One use would be the processing of sensor data, for example, Temperature and humidity for use in the disaggregation.

In General ANN are well suited to solve perception tasks. One possibility is the recognition of complex patterns, visual images of objects, voice recognition, handwritten characters or waveforms of each electronic appliance. The ANN have great potential in terms of the computing time for **pattern recognition**. Classical methods such as Bayesian classifier is far exceeded by ANN.

6.2 Artificial Neural Networks

6.2.1 Perceptron

Analogous to the biological neurons can be set up artificial structures. A simplified model of the structure of an artificial neuron is shown in Fig. 6.3. It is called perceptron. On the right side the compact form of the neuron is shown.

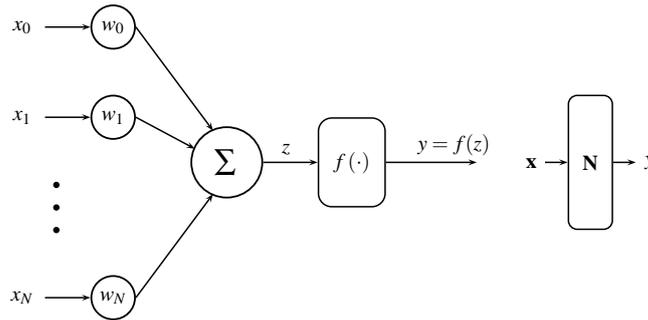


Figure 6.3 – Model of an artificial neuron and its compact structure. This simple form is called perceptron.

The inputs of the neurons are the values x_i which is a vector of length $N + 1$. These values may come from other neurons or from sensors. The first value x_0 can be used to set up the offset. The input values are weighted by the weights w_i and afterward added. The result z is mapped by a non-linear function f to the final result y . In terms of the definitions in chapter 2 the first part is the decision function of the classifier and the function f is the classification function. In the following the model is described in vector notation. The input vector is described by

$$\mathbf{x} = [x_0 \quad x_1 \quad \cdots \quad x_N]^T. \quad (6.1)$$

Analogously the weight vector is given by

$$\mathbf{w} = [w_0, w_1, \cdots, w_N]^T. \quad (6.2)$$

The final output is therefore calculated as

$$y = f(z) = f\left(\sum_{i=0}^N w_i x_i\right) = f(\mathbf{x}^T \mathbf{w}). \quad (6.3)$$

As described before in some cases the input value x_0 is used as an offset. Thus the weight w_0 becomes 1.

6.2.2 Transfer Function

Equation (6.3) applying the non-linear image function is known as the *transfer function*. The easiest case is the step function which fires the neuron when a predefined threshold is exceeded. If the threshold is not exceeded the neuron remains inactive. The disadvantage of using the step function is that the function and its derivative are discontinuous which may result in problems when computing the weights w_i . That is why in most cases the sigmoid function is used (Cichocki and Unbehauen, 1993). This function is determined by the parameters α and β . It is given by

$$y = f(z) = \frac{1}{1 + e^{-\alpha(z-\beta)}}. \quad (6.4)$$

In Fig. 6.4 some examples of the transfer functions are shown.

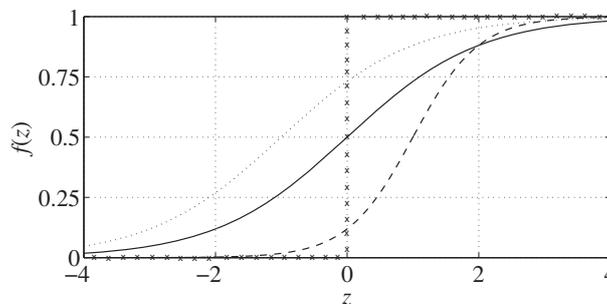


Figure 6.4 – Transfer Functions: Sigmoid Function $\alpha = 1, \beta = 0$ (—); Sigmoid Function $\alpha = 2, \beta = 1$ (- -); Sigmoid Function $\alpha = 1, \beta = -1$ (...); Step Function (xxx).

The gradient of the sigmoid function is defined over the parameter α . The second parameter β shift the sigmoid function. In the most cases the parameters are defined as $\alpha = 1$ and $\beta = 0$ (Kroschel et al., 2011). This equation is also called Fermi-Funtion. The equation 6.4 becomes

$$y = f(z) = \frac{1}{1 + e^{-z}}. \quad (6.5)$$

6.2.3 Structure of Networks

There are several types of artificial neural networks with various structures. The three basic types are separated into the feed forward structure, the feed back structure and the lateral recurrent structure (Cichocki and Unbehauen, 1993). The signal flow of the feed forward structure is from left to right as shown in Fig. 6.5. This structure will be used later. The feed back structure is characterized by the use of feedback paths.

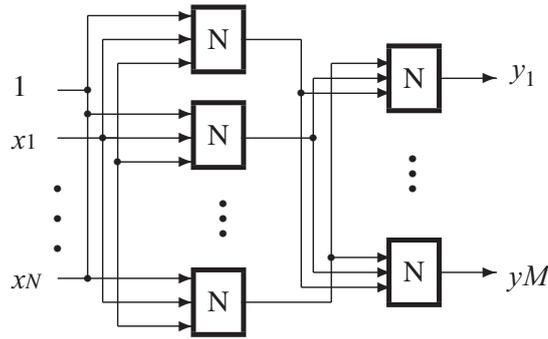


Figure 6.5 – ANN, Feed-Forward-Structure

An even more complex structure is the lateral recurrent structure which makes use of several branches in each of the layers.

The forward structure structure consist of multiple layers. These have several perceptrons. For this reason this type is also called Multi Layer Perceptron (MLP). The structure is shown in Fig. 6.6. It is the most commonly used structure of an ANN. Connections are allowed only between neurons of adjacent layers. There are three distinct layers, namely the input layer, the hidden layers and the output layer.

Now that more neurons are used in multiple layers, the following index is defined. Each neuron is determined by the layer k and the neurons index n . However, it is important to ensure that different layers can have different number of neurons. The activation of a neuron n in the layer k is denoted by $o_n^{<k>}$. The weight of the neuron m in the layer $k - 1$ to the neuron n in the layer k is called w_{nm}^k .

In contrast to the simple perceptron the step function cannot be used as transfer function for the MLP. As already mentioned, this has to do with the differentiation of the function. After the derivation of the learning function, we will see that the transfer function must be continuously differentiable. Therefore a differentiable function defined in the previous chapter must be used. In most cases this is a sigmoid function.

The outputs of the individual neurons can be derived from eq. (6.3). They are calculated as

$$o_n^{<k>} = f \left(\sum_{m=0}^{N_k-1} w_{nm}^{<k>} o_m^{<k-1>} \right) \quad n = 1, \dots, N_k; \quad k = 1, \dots, H \quad (6.6)$$

N_k is the number of neurons in the k -th layer. $o_0^{<k>}$ is analogously to the perceptron the bias and is set to a constant value. The output of the last layer $y = o^{<H>}$ is successively calculated. For this, with the Eq. (6.6) the output is calculated from left to right from the input vector $\mathbf{x} = \mathbf{o}$.

The sigmoid function tends to its asymptotic values. This is not achieved in practice. For this reason, the function values are adjusted. Typical values are 0,1 and 0,9.

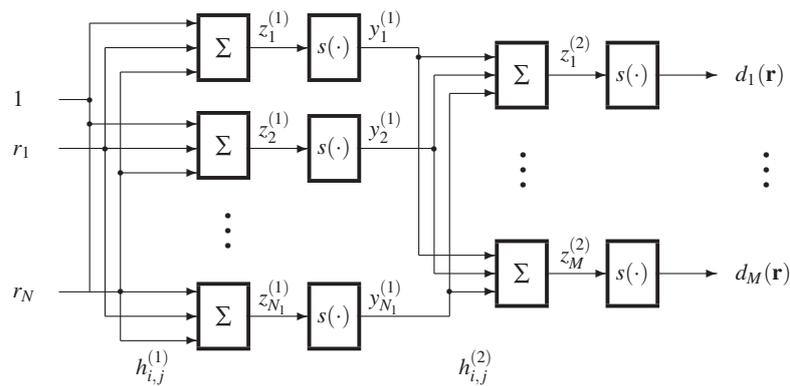


Figure 6.6 – Multilayer Perceptron

6.2.4 Learning Algorithms

The weights w_i plays an essential role for ANN. These parameters are typically implemented not fixed in the neuron. They also cannot be calculated with a closed formula. Instead, they are to learn in an iterative process based on training examples. This goes back to the supervised learning. This then describes the adaptive nature of ANN. In this learning process, the learning rule takes the central position.

In theory of ANN, there are a multitude of different models of learning rules. However, can the most be attributed to a basic hypothesis. This was described by Hebb 1949 (Hebb, 1949) and is called Hebb learning rule. Thereafter, the weights change in proportion to the correlated activity before and after the connection. The weights are calculated as

$$\Delta \mathbf{w} = \mu \cdot \hat{y} \cdot \mathbf{x} = \mu \cdot f(\mathbf{w}^T \mathbf{x}) \cdot \mathbf{x}. \quad (6.7)$$

Here $\hat{y} = f(\mathbf{w}^T \mathbf{x})$ denotes as before the excitation of the neuron, fitting behind the synapses. \mathbf{x} is the stimulus, fitting before the synapses. μ is the size of each learning step. After the trainings phase, the network should have the possibility to bring the write output value to unknown input values. The training phase should seems as follows:

- ▶ apply the input values
- ▶ process the input values (propagation)
- ▶ comparison of target and actual of the resulting output values
- ▶ if it is necessary, minimize the error by adapting the weights (e.g. Perceptron learning rule, backpropagation or LMS)

6.2.4.1 Learning Rule for the Perceptron

The perceptron learning rule represents a supervised learning process. The perceptron will be presented training pairs (x^j, r^j) , $j = 1, \dots, P$. Where x^j is the feature vector and r^j is the associated class. The perceptron learning algorithm is kept simple. From a data set an example (x^j, r^j) is selected randomly. If the classification is correct the algorithm should take the next pair. If not the learning step

$$\Delta w_i = \mu (y_i - \hat{y}_i) \cdot x^j \quad (6.8)$$

should be executed. y_i is 1 when $i = r^j$ or in the other case 0. If the classification for a learning sample j is performed correctly, the weights w_i thus keep unchanged. If an incorrect classification was done the algorithm will forget the wrong answer according to the idea of Hebb's rule (6.7). In return the desired response is learned.

Two criteria can be defined. The first terminated the learning process after a predetermined number of learning steps is reached. The second possibility is to exit the learning process for a defined number of iterations without errors.

6.2.4.2 Learning Rule for the MLP

Until now the learning method was described for one perceptron. Now we describe a method to learn the weights of a multilayer perceptron. With this learning rule, the weights $w_{nm}^{<k>}$ are adjusted. These should be set so that the trainings data are correctly classified or with a minimal classification error. This method is called *backpropagation* (Kroschel et al., 2011). For the derivation of the backpropagation algorithm, the delta rule is extended. For all training data $j = 1, \dots, s$, the summed square classification error is calculated

$$J = \frac{1}{2} \sum_{j=1}^s \sum_{m=1}^{N_H} (y_m^j - \hat{y}_m(\mathbf{x}^j))^2. \quad (6.9)$$

This depends on the weights $w_{nm}^{<k>}$ currently set. J is like a p-dimensional (p is the number of weights of the MLP) area in a p+1 dimensional weight error space. The backpropagation algorithm is now trying to find a local minimum of the error space. For this it uses the gradient descent method. The gradient descent is generally usable, for the iterative process minimizing a function $f(x)$. Here, a correction of the arguments x along the strongest drop of F is performed at each iteration. This is repeated until the gradient of F disappears. This is calculated as

$$\Delta w_{nm}^{<k>} = -\mu \cdot \frac{\partial J}{\partial w_{nm}^{<k>}} \quad (6.10)$$

Before the requirement to determination of $\Delta w_{nm}^{<k>}$ is developed, a simplification should be made. The error, shown in (6.9) is the sum of the contributions of each training example j . A learning method, which computes the gradients on the basis of the total error is called a batch process. However ANN are adapted incrementally. This means that the gradient is calculated for one training example. And after that the adaptation step is performed. This procedure corresponds to the minimization of a error function only dependent on the current training example.

$$J_j = \frac{1}{2} \sum_{i=1}^{N_H} \left(y_i^j - \hat{y}_i(\mathbf{x}^j) \right)^2 \quad (6.11)$$

The requirement for the adaption is given by

$$\Delta w_{nm}^{<k>} = -\mu \cdot \delta_n^{<k>} \cdot o_m^{<k-1>} \quad (6.12)$$

A detailed description of the derivation is published in (Bishop, 2006). The individual parameters are

$$\delta_n^{<k-1>} = f' \left(s_n^{<k-1>} \right) \cdot e_n^{<k-1>} \quad (6.13)$$

$$e_n^{<H>} = \hat{y}_n - y_n \quad (6.14)$$

$$e_n^{<k-1>} = \sum_{m=1}^{N_k} w_{mn}^{<k>} \delta_m^{<k>} \quad k = 1, \dots, H \quad (6.15)$$

where f' is the derivative of the activation function. Now, this shows that f must be differentiable. This becomes clear where the name "backpropagation" is derived. The activities $o_n^{<k>}$ starting from the input to the output layer are calculated as a function of the previous layer $o_m^{<k-1>}$. The errors $\delta_n^{<k>}$ of the neurons are calculated back from the errors of the neurons of the following layer $\delta_m^{<k+1>}$. Proportional to the calculated error then the correction step for each weight $\Delta w_{nm}^{<k>} = -\mu \cdot \delta_n^{<k>} \cdot o_m^{<k-1>}$ is performed on the neurons.

6.2.5 Design of a MLP

When dimensioning the MLP, the number of layers and the number of neurons in a layer must be set. The number of neurons in the input and output layer is defined by the particular requirement. The number of neurons in the input layer is defined by the feature vector. This may consist of signals and parameters. The number of neurons in the output layer is defined by the number of classes.

Whereas the choice of the number of neurons in the hidden layer is not trivial. Research in this area is not yet far. Through experience and experimentation an optimal number of neurons can be determined. Here, two limits must be observed. First, the number

must not be too small. This has the effect that the network is not able to solve the problem. Second, it must be ensured that the generalization is not waived. If too many neurons in the hidden layer, it is possible that the network learn by heart the problem. This can be countered with a reduction in the number of neurons and the validation.

The MLP should have a sufficient number of layers to keep the residual error of the classification small. In the literature it has been proven that a 3-layer perceptron in principle sufficient to model any data / class distribution in p -dimensional space (Bishop, 2006). A proof can be understood clearly by the following consideration. The first layer separates any half-spaces. The second layer summarizes the resulting half-spaces together by conjunct convex areas. The third (output) layer combines the convex regions by disjunction to any concave quantities. Because these models in general MLP consists not of more than three layers.

6.2.6 ADALINE

In the previous section, the simple perceptron and the complex MLP were derived. For both, learning rules have been established which determine the weights. However, it is also possible to determine the weights for an adaptively estimation of the parameters of an unknown system. An **Adaptive Linear Neuron**, called ADALINE, is a combination of a simple neuron with learning rule. It was developed by Widrow and Hoff 1960. For the learning rule the Widrow-Hoff learning rule is used. A combination of more ADALINE in different layers is called MADALINE.

6.2.6.1 Structure of an ADALINE

The ADALINE is a single layer network. The main difference to the perceptron is, that it uses for the learning phase the output of the summation point z_j for the adaptation of the weights. Whereas the perceptron calculates the weights with the signal after the activation function. In Fig. 6.7 an ADALINE is shown.

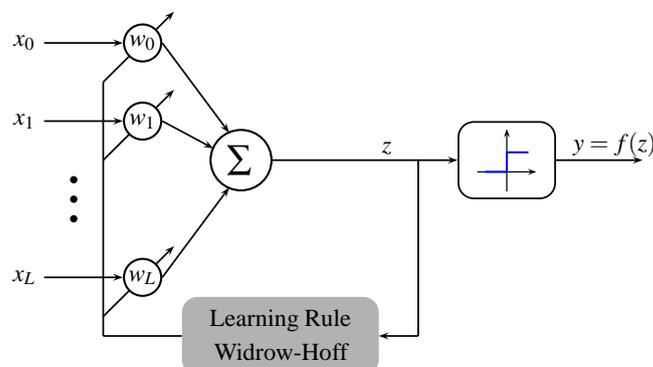


Figure 6.7 – Adaline with Widrow-Hoff learning rule.

6.2.6.2 Learning Rule ADALINE

As described before the Widrow-Hoff learning rule (also known as LMS-algorithm) uses the output signal z for the estimation of the weights. This algorithm is used to set the weights of the ADaptive LInear NEuron (ADALINE). The individual weights of the neurons can be determined in many different ways. The common principle of all known methods is to minimize the error e in order to determine the optimum weight vector \mathbf{w}_{opt} . The block diagram for the computation of the error is shown in Fig. 6.8.

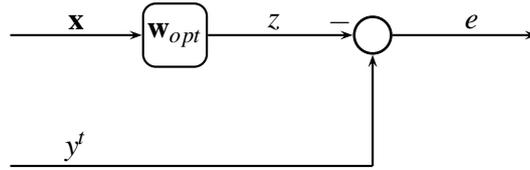


Figure 6.8 – Block diagram for the error e calculation.

For the minimization the same method is used as for the MLP learning rule. The minimization is calculated with the gradient descent method. The optimum is reached when the mean square error becomes minimal. For a weight vector of length two this error is given by

$$\begin{aligned}
 e^2 &= (y^t - z)^2 \\
 e^2 &= (y^t - \mathbf{x}^T \mathbf{w})^2 \\
 e^2 &= (y^t)^2 - 2y^t \mathbf{x}^T \mathbf{w} + \mathbf{w}^T \mathbf{x} \mathbf{x}^T \mathbf{w}.
 \end{aligned} \tag{6.16}$$

Since the vectors \mathbf{x} represent stochastic signals it is useful to use the expectation of the mean square error. The optimum weight vector is computed by finding the minimum of the derivative of equation (6.16). It is given by

$$\mathbf{w}_{opt} = \mathbf{R}^{-1} \mathbf{p}. \tag{6.17}$$

This equation is known as the *Wiener-Hopf Equation* with the auto-correlation matrix $\mathbf{R} = E \{ \mathbf{x} \mathbf{x}^T \}$ of the input signal \mathbf{x} . \mathbf{p} is the cross-correlation vector of the desired value y^t and the input signal \mathbf{x} . A detailed derivation of this principle is given in (Haykin, 2002). The iterative solution becomes:

$$e(n) = y(n) - \mathbf{x}(n)^T \mathbf{w}(n) \tag{6.18}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n) \mathbf{x}(n) \tag{6.19}$$

The step size μ must be estimated. One possibility for the estimation is to use the eigenvalues of the signal. μ can be calculated by the eigenvalues (Kammeyer and Kroschel, 2012)

$$\mu = \frac{2}{\lambda_{max}}. \quad (6.20)$$

λ_{max} is the highest eigenvalue of the input signal. For a practical usage, the step size should be smaller as the quotient of two with the biggest eigenvalue, to get a stable convergence.

6.3 ADALINE for the Classification

In the previous section was extensively discussed the basics of ANN. In this section, a first method is described with which it is possible to classify a group of appliances from the total load profile of residential buildings. For this the household measurements from 4.5.2.3 are used. From Tab. 4.3 we know, that the refrigerators and freezer have a high power consumption in the home. It should be derived features to classify them. For this purpose the ADALINE is used. This is a very simple structure of ANN is selected Benyoucef et al. (2012a). For the decision function the Mean Squared Error (MSE) is used. The classification function consists of a decision threshold maker.

6.3.1 ADALINE for Parameter Estimation

From chapter 2.4 we know, that a periodical signal $p(t)$ can be represented by the Fourier series. It can be divided into a constant component A_0 and the sum of the harmonic oscillations. These oscillations can be divided into a sin- and cos-part. The factors A_n and B_n can be interpreted as the amplitude values of the spectral component. n is an integer and represents the n -th harmonic of the signal.

$$p(t) = A_0 + \sum_{n=1}^N A_n \cos(n \cdot \omega t) + B_n \sin(n \cdot \omega t) \quad (6.21)$$

The ADALINE can be used to estimate the parameters $A_0, A_1, B_1, \dots, A_N$ and B_N . Fig. 6.9 shows the structure of an ADALINE for the harmonics estimation. The signals $x_n(t)$ are the input of the ADALINE. These are harmonic oscillations. $\omega = 2\pi f$ is the frequency of the fundamental wave. They can be summarized in the input vector

$$\mathbf{x} = \begin{pmatrix} 1 \\ x_{A_1}(t) \\ x_{B_1}(t) \\ \vdots \\ x_{A_N}(t) \\ x_{B_N}(t) \end{pmatrix} = \begin{pmatrix} 1 \\ \cos(n \cdot \omega t) \\ \sin(n \cdot \omega t) \\ \vdots \\ \cos(n \cdot \omega t) \\ \sin(n \cdot \omega t) \end{pmatrix} ; n = 1, \dots, N \quad (6.22)$$

Now we are looking for the weight vector \mathbf{w} . This can be interpreted as the amplitude values of the spectral components. The weight vector is

$$\mathbf{w} = [A_0 \ A_1 \ B_1 \ \cdots \ A_N \ B_N]^T \quad (6.23)$$

$p(t)$ is the desired output. The estimated signal is $p_{est}(t) = \mathbf{x}^T \mathbf{w}$. With the Widrow-Hoff rule (cf. (6.19)) the error e will be minimized and the weights are calculated.

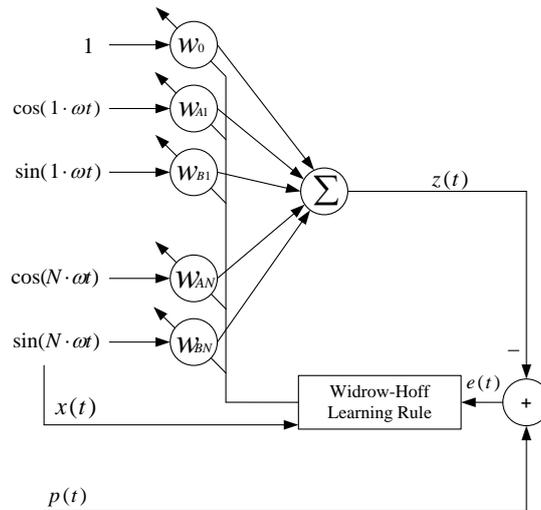


Figure 6.9 – ADALINE for the harmonics estimation.

6.3.2 Simulation

The ADALINE was simulated. For the input signal $p(t)$ the fundamental wave was superimposed with the 2, 3 and 4 harmonic. sin- and cos-parts were used. Therefore the desired weight vector should be $\mathbf{w} = [A_0 = 0 \ A_1 = 0 \ B_1 = 2 \ A_2 = 0 \ B_2 = 3 \ A_3 = 0 \ B_3 = 1 \ A_4 = 5]$. The frequency of the signals is $f = 50$ Hz.

$$p(t) = 2 \cdot \sin(\omega t) + 3 \cdot \sin(2 \cdot \omega t) + \sin(3 \cdot \omega t) + 5 \cdot \cos(4 \cdot \omega t)$$

Eq. (6.19) was used to calculate the weights of the ADALINE. The step size was estimated with the eigenvalue of the signal. Therefore eq. (6.20) was used.

$$\mu = \frac{2}{\lambda_{max}} = \frac{2}{139,6} = 0,0144$$

Fig. 6.10 shows the simulation of the parameter estimation. In the first row the input signal $p(t)$ is shown. In the second row the error e and in the last row the harmonics are shown.

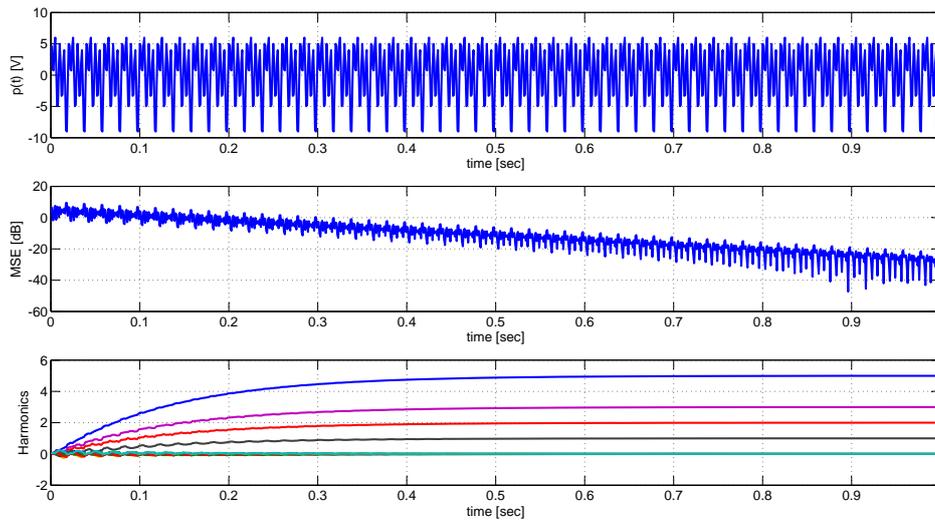


Figure 6.10 – Results for the parameter estimation. In the first row the desired signal is shown. In the second row it is shown that the error is decreasing so the algorithm converges. In the last row the convergence of the harmonic values are shown.

The algorithm converges. The termination criterion of the learning process is reached when the MSE falls below a defined threshold $e \leq e_{min}$. For this simulation the learning process should be stopped when the error falls below 0,001 ($MSE = -30dB = 10 \log(0,001)$). Then the weight can be read. In Fig. 6.11 the weights after the termination criterion are presented. The weight vector consists of the amplitude values of the harmonics.

$$\mathbf{w} = [0 \ 0 \ 2 \ 0 \ 3 \ 0 \ 1 \ 5]^T$$

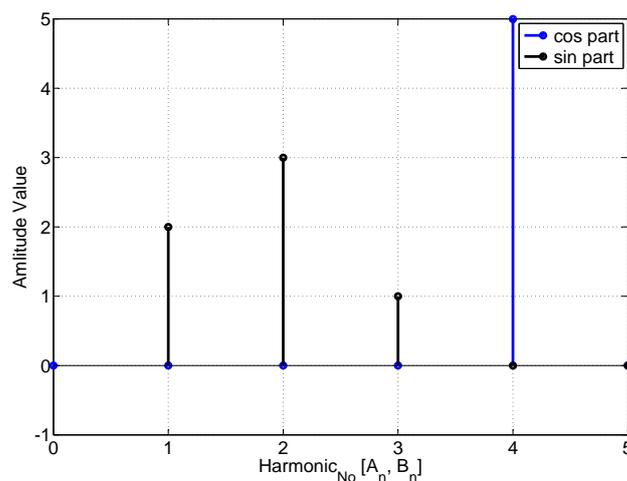


Figure 6.11 – Weights of the ADALINE after the learning phase. The weights correspond to the harmonics of the desired input signal $p(t)$.

6.3.3 Results

In the simulation, it was shown that the ADALINE can estimate the harmonics of a periodic signal. These are used as features for the classification. In following this approach should be applied to real measurements. The data set of the home measurements includes 50 recorded days (cf. 4.5.2.3). Fig. 6.12 shows the power consumption of approximately 43 hours.

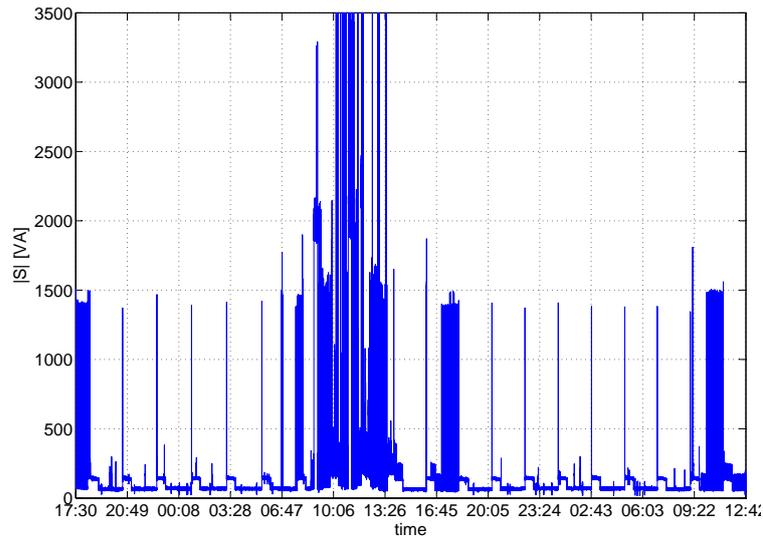


Figure 6.12 – Power consumption of a household. During the night phase the switching events of the refrigerator are visible.

The ADALINE is to be used to classify the refrigerator from the total load profile. The periodic cycles of the refrigerator are clearly visible in the night phase. In Fig. 6.13, the switching on phases of the first two seconds of the refrigerator, is shown. As a comparison, the algorithm from Hart was implemented. This waits 2 s until it reaches the steady state. Then it executes on the basis of load change the classification (Hart, 1992).

The approach based on the ADALINE should improve the classification by decreasing the time period. The profile will have a short time to extract the features. The vector which is used for the feature extraction includes only 25 samples. This is similar to 0.5 s. So the new approach based on ANN is of factor 4 better as the approach of Hart. This has the main advantage, that the minimum time between two switching events of appliances of decreasing.

Fig. 6.14 shows the learning phase of the ADALINE. In the first row the switching on cycles of the refrigerator is shown. The error is shown in the second row. The last row presents the estimated weights. By using eq. (6.20) and an eigenvalue of the signal from 10,708 the maximum step size becomes $\mu = 0,0934$. In this practical example a step size with $\mu = 0,04$ was used. The error is already covered by one switching cycle. After 2 cycles the weights are final estimated.

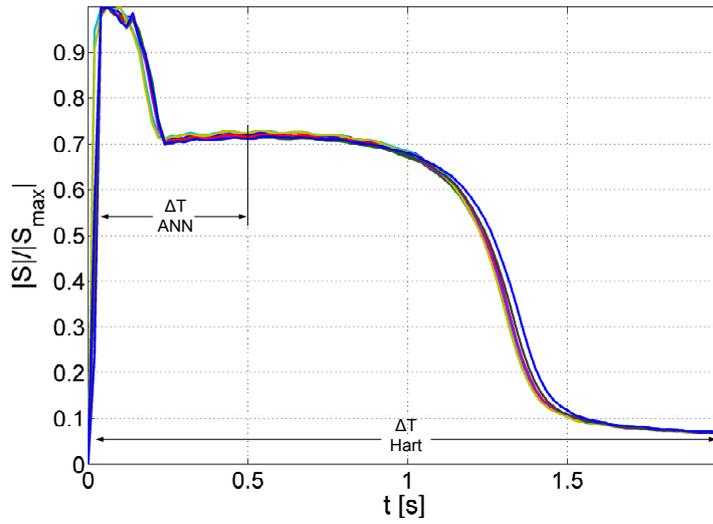


Figure 6.13 – Different switching on phases of a refrigerator, with the delta time of the two approaches.

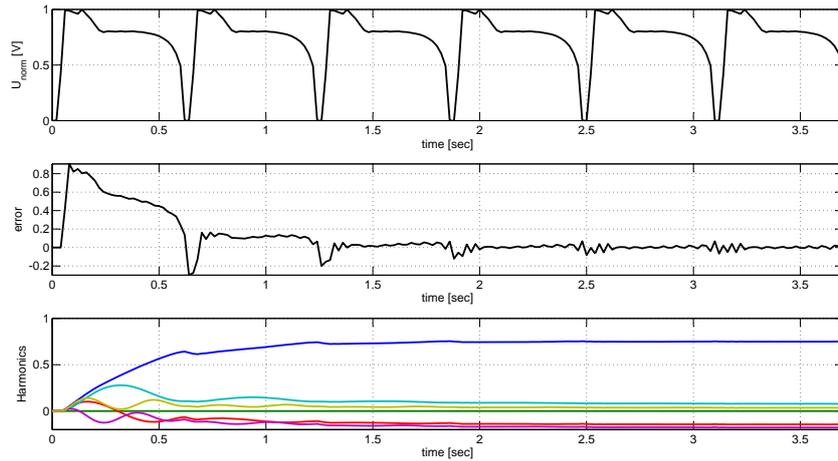


Figure 6.14 – Results for the feature extraction with an ADALINE. With real switching cycles of the refrigerator the true values of the vector \mathbf{w} are calculated.

For the decision function the MSE should be used. It is calculated as

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{w}_n - w_n)^2 \quad (6.24)$$

$$MSE_{dB} = 10 \cdot \log_{10} \left(\frac{1}{N} \sum_{n=1}^N (\hat{w}_n - w_n)^2 \right) \quad (6.25)$$

w_n is our known feature vector from the refrigerator. \hat{w}_n is the vector which includes the harmonics of the current switching cycle. The MSE calculates the sum of the

variance and the squared bias of the harmonic estimation. The MSE is the input signal of the decision threshold maker (cf. classification function event detector 5.4.5). This was used for the classification. If the MSE fall below a predefined threshold a similar appliance like the refrigerator is classified. Otherwise an other appliance was switched on.

Altogether 1235 switching on events of the refrigerator were measured in the recorded time. The approach of Hart and the approach of ADALINE were implemented and tested. The classification algorithm based on Harts approach (Hart, 1992) has classified 1080 (87.5%) of the switching on events. The algorithm based on the ADALINE approach has classified 1162 (94.1%) of the switching on events. The results are presented in Tab. 6.1.

Approach	N_e	TP	FP	TPR	PPV
Hart	1235	1080	262	87.5 %	80.47 %
ADALINE	1235	1162	13	94.1 %	98.89 %

Table 6.1 – Results for the classification. Two algorithms were analyzed. The results of the approach of Hart are shown in the first line. The results using the ADALINE approach are shown in the second line.

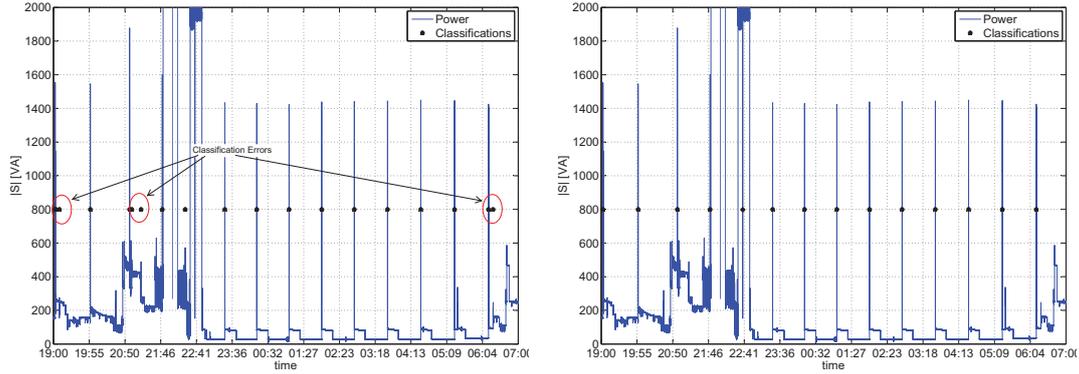
Both approaches have a good classification rate, whereas the approach based on ADALINE has 6.6% more switching events classified. Both approaches have also classification errors. Thus the approach of Hart has 262 FPs, because there are appliances with the same power consumption as the refrigerator. The ADALINE approach has 13 FPs, this depends on noise in the power signal. Thus, for the classification errors, the ADALINE approach is much better as the approach of Hart Bier et al. (2012).

In Fig. 6.15 is an example for the results of the classification for a real measurement. In Fig. 6.15(a) the results of the classification based on the approach of Hart are shown. In Fig. 6.15(b) the results of the classification based on the approach of ADALINE are shown. Between the time, where many appliances are switched on (approx. 19:00-23:00 and on the other day between 6:00-7:00) the approach of Hart has some error classifications. Whereas the approach based on ADALINE has no error classification.

6.4 MLP for the Classification

In the previous section an ANN were used to estimate parameters of the transient response of individual appliances. As ANN an ADALINE was used. This can solve a linear problem. For more complex problems a non-linear structure must be used. In theory part, it has been described that a MLP satisfy this property. In this section, a MLP should be used to classify different appliances.

Two different feature sets are described, which can be used for the classification. After that the results are compared and discussed.



(a) Classification based on the approach of Hart. All TP in the sense of classifying the refrigerator are reached. But also some FP are classified. (b) Classification based on the approach of ADA-LINE. All TP in the sense of classifying the refrigerator are reached. And there are no FP.

Figure 6.15 – Example of the results for the classification. Approximately 12 h are presented of the load profile of the measured household.

6.4.1 Features for the Classification

For the classification different features can be used. In that thesis two types of features are presented and compared. The first one are the power signals P , Q and S . From the switching on event ev of the appliance a defined sequence of samples n of the signals are used for the classification. The input vector \mathbf{x}_a of an appliance a for the MLP is

$$\mathbf{x}_a = [P(ev_a) \quad P(ev_a + 1) \quad \cdots \quad P(ev_a + n - 1)]^T. \quad (6.26)$$

The second features are the output values from the STFT of the switching on phase of the power signal. From the power sequence of 20 samples, the STFT is calculated. If the window length is diversifying, the length of the STFT is varies. The results of the STFT are squared, so we receive the power density values of the signal. With the time discrete function of eq. (5.30), the input signal of the MLP becomes a Matrix

$$\mathbf{x}_a = \begin{bmatrix} \sum_{n=-\infty}^{\infty} p(n)w(n-1)e^{-j2\pi fn} \\ \sum_{n=-\infty}^{\infty} p(n)w(n-2)e^{-j2\pi fn} \\ \cdots \\ \sum_{n=-\infty}^{\infty} p(n)w(n-m)e^{-j2\pi fn} \end{bmatrix}. \quad (6.27)$$

m is the number of parts, where the input sequence is divided.

6.4.1.1 Train the MLP

From chapter 2 we know that a learning process of a MLP can be divided into the training, the validation and the test phase. Therefore the data-set is divided into three parts. In practical use the main part of the data-set is used for training. For validation and test the smaller rest is used. The signal flow diagram of an ANN is shown in Fig. 6.16.

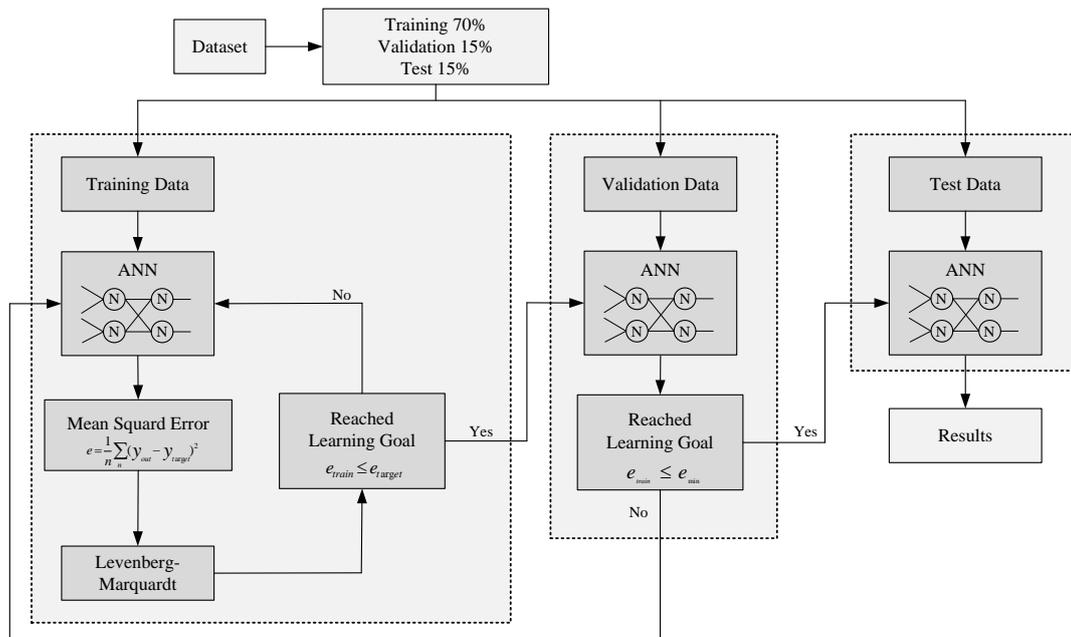


Figure 6.16 – Diagram of the signal flow from the training, validation and test of the ANN.

The weights are calculated over the error from the outputs, where the input values are the test data. During the test phase, with the validation data, the results are verified with known outputs. After a predefined error value e_{min} is decreased, the network is tested with the test data, where the outputs are unknown.

6.4.2 Results of the Classification Method

For the classification a MLP is used. This has a feed forward structure. It has three layers (this number is sufficient in most cases). The input layer consists of the different feature vectors \mathbf{x}_a . The number of neuron for the hidden layer was determined iteratively. The optimal number of neurons in sense of optimization the TPR was 10. The number of neurons in the output layer depends from the number of classes. Since we have 10 appliances, the network has 10 output neurons. Every output images one of the 10 appliances. The transfer function is a non-linear sigmoid function with $\alpha = 1$. This is necessary for the derivation of the learning rule.

Every feature set was divided into three parts. 70% of the data are used to train the network, 15% are used for validation the network and the other 15% are used to test

the network with an independent data set. For the training of the network and to find the optimal weights, the back propagation learning rule was used. To solve the numerical optimization of the learning rule, the Levenberg-Marquardt-Algorithm was used.

For the analyze of the approach the measurements described in chapter 4.5.2.2 were used. Altogether 10.000 patterns from 10 different appliances were classified. For a better presentation of the results the classification rates are presented in two tables. Tab. 6.2 shows the TPR for the first 5 appliances. Tab. 6.3 shows the TPR for the last 5 appliances. The first three rows of the results consist of the feature with the signal power. The best results were reached with the active power P . The length of the sequence n is diversified. In the first row 10 samples (200 ms) are used as input vector. Then in the second row 15 samples (300 ms) and in the third row 20 samples (400 ms) are used for the input vector. It is well shown that increasing the number of sample n , the TPR increase. This comes from the different shape of the switching on cycle of the appliances. The longer the input vector for the ANN is, the better can correlate and distinguish between the trained appliances. For a real-time system, the number n should be kept as small as possible. There are two reasons, firstly, the computational effort should be as low as possible and on the other the time in which two appliances can be switched on should also be small. With $n = 20$ samples it is a good compromise, because the ANN has a TPR over all appliances of 97.1 % and the time in which no two appliances can be switched is less than 400 ms.

Features	Water Heater	Handheld Mixer	Freezer	Hair Dryer 1	Mixer
P, $n = 10$	84.1 %	85.7 %	68.6 %	95.8 %	98.6 %
P, $n = 15$	89.7 %	93.8 %	82.9 %	97.4 %	98.3 %
P, $n = 20$	98.6 %	98.6 %	92.7 %	98.6 %	98.6 %
STFT, $m = 2$	94.2 %	98.5 %	88.8 %	97.6 %	100 %
STFT, $m = 3$	76.0 %	93.0 %	90.4 %	98.4 %	98.3 %
STFT, $m = 4$	94.1 %	95.0 %	95.2 %	92.0 %	100.0 %

Table 6.2 – First part of the classification results for the measurements using a MLP. Two features are compared.

Features	Lamp	Heater	Energy Lamp	Hair Dryer 2	Radio
P, $n = 10$	92.0 %	90.6 %	90.7 %	80.5 %	90.9 %
P, $n = 15$	100 %	89.4 %	96.6 %	95.5 %	97.1 %
P, $n = 20$	98.8 %	94.3 %	96.0 %	96.1 %	100 %
STFT, $m = 2$	0.0 %	94.9 %	98.6 %	88.5 %	3.8 %
STFT, $m = 3$	49.3 %	81.8 %	100 %	77.6 %	0.0 %
STFT, $m = 4$	0.0 %	92.0 %	100 %	80.0 %	3.4 %

Table 6.3 – TPR of the classification of the appliances with ANN and different input features. Two features are compared.

In row 4 until 6 the TPR of the 10 appliances with the input vector of the second feature are shown. In these three rows different resolutions of the spectrum are used as input

vector. The signal was divided into $m = 2, 3$ and 4 parts. For some appliances this feature is good for the classification. For appliances with a smaller power consumption like radio or lamp, the feature is insufficient. Because the harmonics lost in the noise.

If we compare the two results from the feature power and spectrum it is shown, that the two approaches has different TPR. To classify some appliances the first feature has better results and to classify the other appliances the second feature has better results. Overall a combination of both features should be used to increase the TPR of the classification the appliances.

6.5 Conclusion

In this chapter approaches for the classification of appliances in the load profile have been described. It is the last block of the processing chain, which has been analyzed in this thesis. After the conditioning of the signals and the event detection the classifier attaches the detected appliance to a class.

There are different kinds of classifiers. Since in this thesis a data set with reference values was available, a supervised method was chosen. Based on preliminary studies and the state of art an ANN was used. It was given a general overview of neural networks. This includes the idea of the technical implementation of biological neurons. Different applications of ANN were presented

There exist different structures of ANN. The simplest one is the perceptron. This was described mathematically. Different transfer functions which fires the neuron were presented. It is possible to link several neurons together. Another structure is the MLP. This consists of several layers with different number of perceptrons.

Essential components of the ANN are the weights. To determine this from the data set, for the perceptron and the MLP different learn algorithms were derived. The learning rule for the perceptron is based on the Hebb rule. Whereas the learning rule for MLP based on the gradient descent method. This method leads to backpropagation algorithm. As the final structure, an ANN has been described which can be used for parameter estimation. It adapts its weights. For this reason it is called adaptive linear neuron ADALINE.

After describing the theoretical basics on ANN, 2 classification approaches have been described. The first approach uses a ADALINE structure to estimate the harmonics of the transient response of appliances. The group of freezer and refrigerators should be classified. For the decision function the MSE was used. The classification function was a threshold decision maker. The approach was analyzed on 1235 switching events. It was compared with the approach of Hart. It was shown, that the approach using the ADALINE can improve the TPR by 6.6%. Even better is the improvement of the PPV. The approach with the ADALINE has improved the PPV by 18.4% in comparison to Hart.

For the second classification method, a MLP has been designed. The measurements from 4.5.2.2 were used for the training, validation and test phase. Overall 10.000 patterns of 10 different appliances were used. Two feature vectors were used. The first

is based on the power signals. The second feature vector contains the harmonic of the switching on phase calculated with the STFT. It has been shown that it is possible to classify very effective individual appliances with this approach. A combination of the individual features results in a higher detection rate.

7 Conclusion and Perspectives

The issue of energy monitoring has increased significantly in popularity in recent years. This makes it possible to use energy more efficiently. With new approaches from disaggregation the energy consumption should be reduced in residential buildings. One way to do this is NALM. This has some advantages compared to an invasive system. One important that the wiring effort is lower and thereby reduces the cost.

This thesis deals with problems in the area NALM. It is part of the research project *Smart Metering* at the HFU. In this thesis three topics will be treated in more detail.

1. Measurement system and analyze of signals for NALM
2. Event detection
3. Classification

At the beginning of the thesis, general expressions were defined in chapter 2 in the field of pattern recognition. Classifiers were presented. Since the measurements in this thesis, included the input and the target function respectively, a supervised method was selected. For the statistical analysis of algorithms, it is essential to define quality criteria. Using these criteria, the individual approaches can be compared. At the end of the chapter a method was presented, which enables an analysis of the signals in frequency domain. Based on the FT, the STFT was derived. This allows to investigate non-stationary signals. A disadvantage of the STFT can be avoided with the S-Transformation. This has also been described and may be used for later work.

The theme NALM has its beginnings in the 80's of last century. Since that time, different approaches have been developed by several research groups worldwide. Chapter 3 gives an overview of different approaches. They can be classic divided into the steady state analysis and the transient state analysis. New methods combine both approaches or focusing on new analyzing methods. Therefore, the classical division has been extended in this thesis.

For a statistical analysis of disaggregation algorithms it is necessary to have a multitude of measurements. This must have good documentation of the time stamps from the switching events. Since the beginning of this thesis was no suitable data-set exists, an own data-set should be created (cf. chapter 4). To record a data-set, an own measurement system was developed. This can measures one- and three-phases of the currents and voltages. The measuring system was statistically analyzed. Thus it was demonstrated that there is no disturbing influence on the measurements. The transfer characteristic corresponded to the specifications. The SNR is negligible and the harmonics have no significant influence on the measurements. After analyze of the measurement system first measurements were performed. The behavior of the appliances was investigated. For this the transient response was compared. The individual

powers P , Q , and S were investigated. By the end of the thesis are in the database, three types of measurements.

- ▶ 400 individual measurements of appliances from 15 residential buildings.
- ▶ 10000 switching events from 10 different appliances. These differ in the variation of their power consumption.
- ▶ 50 days residential buildings measurements of three phases.

The event detection was the second great part which has been worked in this thesis. Switching on- and off-events of appliances lead to abrupt changes in the signal curve. This event is very important for the disaggregation. At these points the later classification will be divided the individual appliances in their classes. In chapter 5 fundamental terms for the event detection are described. An overview of methods for event detection was given. In this thesis three methods for event detection were investigated in detail. The first uses a high-pass filter structure. Based on the gradient operator and the Laplace operator different filter masks were derived. The second method also uses a filter structure. It is based on the mathematical description of the problem and was described by Perona. The filter masks are complex and consist of a double derived Gaussian function and their Hilbert transform. The third method is based on the analysis of the signal in the frequency domain. For this purpose, the STFT was used. This makes it possible to analyze time-varying signals. All three methods have been simulated and implemented. The methods were compared with the well-known approach of Hart. It is shown that the Perona filter has a significant improvement over the approach of Hart. The noise reduction is much better and therefore not as many FP are detected. As a result it has been shown in this thesis that the Perona filter for event detection provides the best detection rates. For later implementation in a real-time smart meter system, this filter should be used.

The last great chapter, which was processed in this thesis deals with the classification of appliances. After the detection of a switching on event the classifier perform a classification of the appliances at this point. In this work, a classification method was sought which is in the group of supervised learning methods. From the state of art and own preliminary studies, a classifier was chosen based on artificial neural networks. Their structures allow a very fast parallel data processing. In chapter 6 a detailed description of artificial neural networks is presented. There were presented two methods for classification. The first one based on an ADALINE. This model estimates the parameters of the switching on event of a refrigerator. This is passed as a feature vector to the decision function. For this MSE was used. With a decision threshold maker is then classified whether the device is a refrigerator or not. The method was simulated and implemented. As a reference algorithm, the algorithm of Hart has been used. The results show that the method based on ADALINE improves the detection rates. The TPR increases by 6.6%. However, much better is the increase of the PPV. This was increased by the new method by 18.4%. It has been shown that this classification method achieves good improvement over known methods. The second classification method is based on the use of a multilayer perceptron. For the verification of that approach, a data-set was used consist of 10.000 measurements with 10 different appliances. The MLP was trained with this data-set. Two different feature vectors were used. One is based on the time signals of the power, the other uses the harmonics

of the transient response calculated with the STFT. The results show that the MLP is able to classify the most appliances. With very good TPR of about 95 % it shows that the feature P achieves the best rates. This chapter has shown that if a sufficient data set is to available simple appliances can be detected well.

In this thesis the fundamental blocks MS, ED and CL have been processed. However, further work is required for the development of a practical monitoring system which can indirectly contribute to the reduction of electricity consumption. These have already been described in the research project *Smart Metering*. In a following research project which begins at the end of this year, two other PhD students should working on the development of a deploy-able monitoring system. These should work on other problems in the area of NALM. In the following an overview is shown of possible further work. But this does not claim to completeness.

1. *Classification of Complex Loads.*

Intended to classify complex loads, the use of other feature vectors are examined. Promising is the use of the S-Transform (Jimenez et al., 2014). It can be used for the investigation of the harmonics of the current (Lin and Tsai, 2014). As already described in the thesis, the S-Transform uses a frequency adjusted window. On the signal level of voltage and current and other methods may be used. One example is the use of V-I trajectories, described in (Hassan et al., 2014).

2. *Classification using Secondary Information*

For the problems of tracking and classification secondary information can be used. This may include a-priori information. In the classification of the appliance, the location and time of day was called as a criterion. Next, environmental factors such as temperature or humidity may be used (Kazakidis et al., 2012).

3. *Modeling and Combination of Complex Loads*

The results in this thesis have been shown that simple on-off appliances with different power consumptions can be well classified. From the preliminary investigations it is known that automates in the ground consist of on-off appliances. For example, a washing machine consists of a heating element and a motor. For this purpose, probabilistic models of the switching times of the appliances should be created (Wichakool et al., 2014).

4. *Tracking of the Power Consumption.*

After the most alliances can be classified in the household, the power consumption must be calculated. For this purpose, a study of method is necessary, which assign a switching off event to a switching on event of an appliance. Thus, the operating time can be determined which is used for the calculation of the power consumption (Jung et al., 2012; Xu and Dong, 2012).

5. *Visualization of the Results.*

An effective energy saving is achieved only through a change in the behavior of consumers. The monitoring system should show the consumers ways to save energy and use energy more efficiently. For this purpose a transparent visualization of the results is important. The results in this work and which will be attained in the following may not only be presented to the consumer. It must be made concrete proposals how energy can be saved (Goodwin et al., 2013).

One possibility would be to identify inefficient appliances and in a second step to present proposals for other energy-saving types.

Appendix

A Measurements and Measurement System

Contents

A.1	Signals for NALM	128
A.2	Measurements in the Laboratory	129
A.3	Measurements in Residential Buildings	131
A.4	Harmonics of the Current	132
A.5	Transient Responses of Appliances	134

A.1 Signals for NALM

The following two tables list the most important signals for NALM are summarized. Tab. A.1 shows the signal for a resistance R and the impedance Z . In Tab. A.2 the signals for the inductance L and capacitance C is shown.

Description	Resistance R	Impedance Z
fundamental law	$u = R i$	
ohms law		
absolute value	$I = GU = U/R$	$I = YU = U/Z$
complex values	$\underline{I} = \underline{GU} = \underline{U}/R$	$\underline{I} = \underline{YU} = \underline{U}/\underline{Z}$
resistance	$R = U/I$	$Z = U/I$
complex		$\underline{Z} = \underline{U}/\underline{I} = R + jX$
admittance	$G = I/U = 1/R$	$Y = I/U = 1/Z$
complex		$\underline{Y} = \underline{I}/\underline{U} = 1/\underline{Z} = G + jB$
phase angle	$\varphi = 0^\circ$	$\varphi = \arctan(Q/P)$
active factor	$\cos(\varphi) = 1$	$\cos(\varphi) = P/S = R/Z = G/Y$
reactive factor	$\sin(\varphi) = 0$	$\sin(\varphi) = Q/S = X/Z = -B/Y$
active power	$P = UI$	$P = UI \cos(\varphi) = S \cos(\varphi)$
reactive power	$Q = 0$	$Q = UI \sin(\varphi) = S \sin(\varphi)$

Table A.1 – Overview of the most important signals of a resistance R and an impedance Z for passive sinusoidal current for a two-terminal network.

Description	Inductance L	Capacitance C
fundamental law	$u = L di/dt$	$u = 1/C \int i dt$
ohms law		
absolute value	$I = -B_L U = U/X_L$	$I = B_C U = -U/X_C$
complex	$\underline{I} = -jB_L \underline{U} = \underline{U}/jX_L$	$\underline{I} = jB_C \underline{U} = -\underline{U}/jX_C$
resistance	$X_L = \omega L = U/I$	$X_C = \frac{-1}{\omega C} = \frac{-U}{I}$
complex	$j X_L = j\omega L$	$j X_C = \frac{-1}{j \omega C}$
admittance	$B_L = -1/(\omega L)$	$B_C = \omega C$
complex	$j B_L = 1/(j\omega L)$	$j B_C = j \omega C$
phase angle	$\varphi = 90^\circ$	$\varphi = -90^\circ$
active factor	$\cos(\varphi) = 0$	$\cos(\varphi) = 0$
reactive factor	$\sin(\varphi) = 1$	$\sin(\varphi) = -1$
active power	$P = 0$	$P = 0$
reactive power	$Q = UI$	$Q = -UI$

Table A.2 – Overview of the most important signals of an inductance L and a capacitance C for a passive sinusoidal current two-terminal network.

A.2 Measurements in the Laboratory

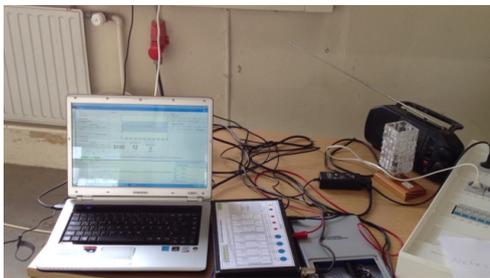
As described in chapter 4.5.2.2, measurements were performed in the laboratory. This has the advantage that in a short time, a large data-set can be generated. These provided a good documentation of the switching events. The measurements were used for the development of the event detector and the classifier. They can also be used for the development of further methods.

The individual measurements were stored in several files. An overview of the current measurements is shown in Tab. A.3. The measurements with the 10.000 events, which were used for the development of the ED and CI having the index $No = 1.x \ x = 1, \dots, 10$. There are still carried out further measurements, which vary in the number of appliances and the switching events.

No	Name of the Measurement	Appliances	Events
1.1	CI-Szn_1-Measurement_1	10	1000 (500 On / 500 Off)
1.2	CI-Szn_1-Measurement_2	10	1000 (500 On / 500 Off)
1.3	CI-Szn_1-Measurement_3	10	1000 (500 On / 500 Off)
1.4	CI-Szn_1-Measurement_4	10	1000 (500 On / 500 Off)
1.5	CI-Szn_1-Measurement_5	10	1000 (500 On / 500 Off)
1.6	CI-Szn_1-Measurement_6	10	1000 (500 On / 500 Off)
1.7	CI-Szn_1-Measurement_7	10	1000 (500 On / 500 Off)
1.8	CI-Szn_1-Measurement_8	10	1000 (500 On / 500 Off)
1.9	CI-Szn_1-Measurement_9	10	1000 (500 On / 500 Off)
1.10	CI-Szn_1-Measurement_10	10	1000(500 On / 500 Off)
2.1	ED-Szn_1-Measurement_1	fan heater	120 (60 On / 60 Off)
2.2	ED-Szn_1-Measurement_2	fan heater	360 (180 On / 180 Off)
2.3	ED-Szn_1-Measurement_3	fan heater	200 (100 On / 100 Off)
2.4	ED-Szn_1-Measurement_4	Lamp	360 (180 On / 180 Off)
2.5	ED-Szn_1-Measurement_5	Lamp	120 (60 On / 60 Off)
2.6	ED-Szn_1-Measurement_6	Lamp	120 (60 On / 60 Off)

Table A.3 – Overview of the current available measurements from the laboratory simulation.

The measurements were done with the own measurement system. Some pictures of the measurements are shown in Fig. A.1.



(a) Pc with ADC and measurement box 1



(b) SDB 1 in the left and some different appliances, e.g. freezer, lamp and oven



(c) the whole system with different appliances



(d) the whole system with different appliances

Figure A.1 – Pictures of the measurement system. Using in the laboratory environment to create a big data-set of measurements in a short time.

A.3 Measurements in Residential Buildings

As described in chapter 4.5.2.3, measurements were performed in two residential buildings. The individual measurements were stored in several files. An overview of the current measurements is shown in Tab. A.4. From a refrigerator 1235 switching events exists. From the freezer which has a similar transient response as the refrigerator 1145 events exists. The two appliance are plugged on different phases. From this reason there have no influence to each other.

No	Name of the Measurement	Freezer	Events	Refrigerator	Events
1-9	Home_Measurement_1_x	L3	160	L1	190
10-56	Home_Measurement_2_x	L3	985	L1	1045

Table A.4 – Overview of the current available measurements from the households measurements.

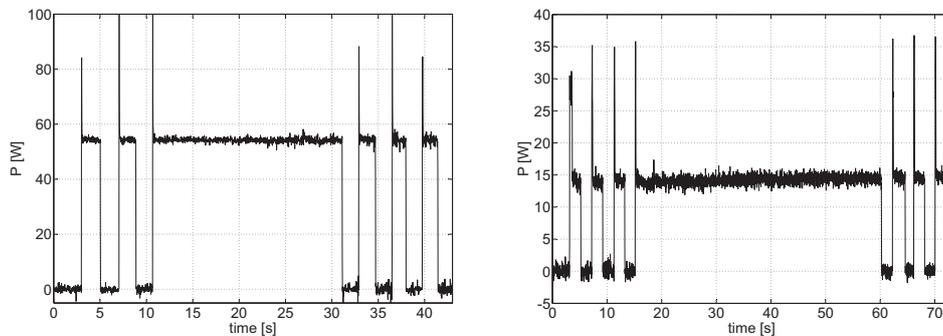


Figure A.2 – Measurement in a residential building. The sensors are plugged directly to the household entry point. The voltage and current of three phases were measured.

A.4 Harmonics of the Current

In chapter 4.6 the behavior of the measured appliances were analyzed. The signals P , Q and S were used. It was shown the transient response as well as the individual distributions of P , Q and S . As previously described these signals are the basis for the development of the algorithms in this thesis. $u(t)$ and $i(t)$ should not be used. For the reason that the high sampling rate is not achieved in today's smart meters. But from chapter 3 we also known that distortions in the current can be analyzed. For the sake of completeness a short analyzing of the harmonics of the current is described here. Based on this, new method can be developed in further work.

As an example, two different kinds of lamps are analyzed. First, we look at the instantaneous power. Fig. A.3(a) shows the instantaneous power of the incandescent lamp with 60 W. The instantaneous power of the energy saving lamp with 15 W is shown in Fig. A.3(b).

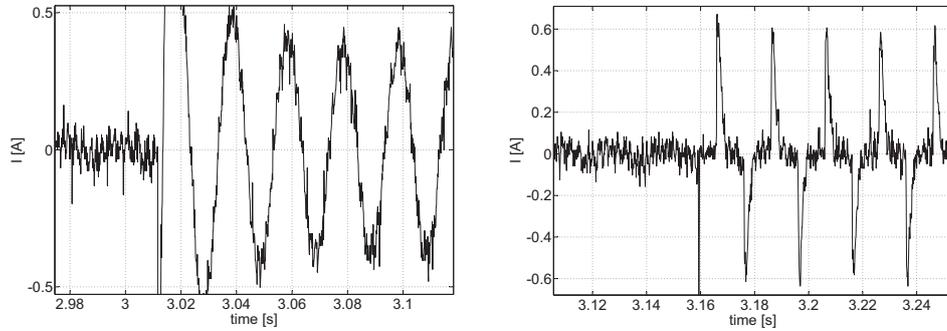


(a) Instantaneous power of incandescent lamp 60 W. (b) Instantaneous power of energy saving lamp 15 W.

Figure A.3 – Instantaneous power of a incandescent lamp and an energy saving lamp.

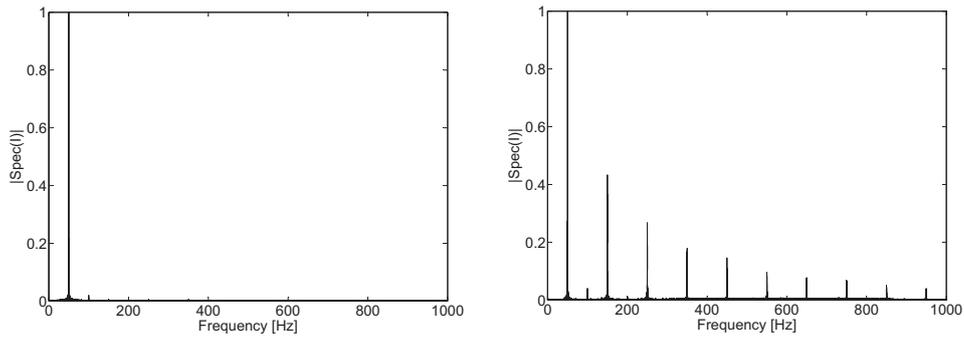
Based on the instantaneous power it is not directly possible to distinguish between these two appliances. For this purpose it is necessary to use other features like the harmonics of the current. A plot of the current signal of the two lamps is shown in Figures A.4. The current profile of the incandescent lamp is sinusoidal whereas the energy saving lamp has the profile of a switching power supply with short peaks. From the current profile can be extracted different characteristics. Once the transient characteristic. Furthermore, the typical time curve.

Another criterion for detecting active states of loads, is the analysis of the harmonics of the current in steady state. This is accomplished by transforming the current profile into the frequency domain. This can be done by using the STFT or if a variable window is necessary by using the ST. Fig. A.5 shows the harmonics of the current for the two lamps. In Fig. A.5(b), the energy-saving lamp clearly shows more significant harmonics than the incandescent lamp in Fig. A.5(a). This is because the energy saving lamp operates on a switched power supply, while the normal lamp only consumes power on the fundamental wave. It is shown here that the harmonics can be used for classification of appliances.



(a) Incandescent lamp with a power consumption of 60 W. (b) Energy saving lamp with a power consumption of 15 W.

Figure A.4 – Time series of the current for one class of appliances.



(a) Incandescent lamp with a power consumption of 60 W. (b) Energy saving lamp with a power consumption of 15 W.

Figure A.5 – Harmonics of current signal for one class of appliances.

A.5 Transient Responses of Appliances

In the following some examples of transient responses of different appliances in the household are shown.

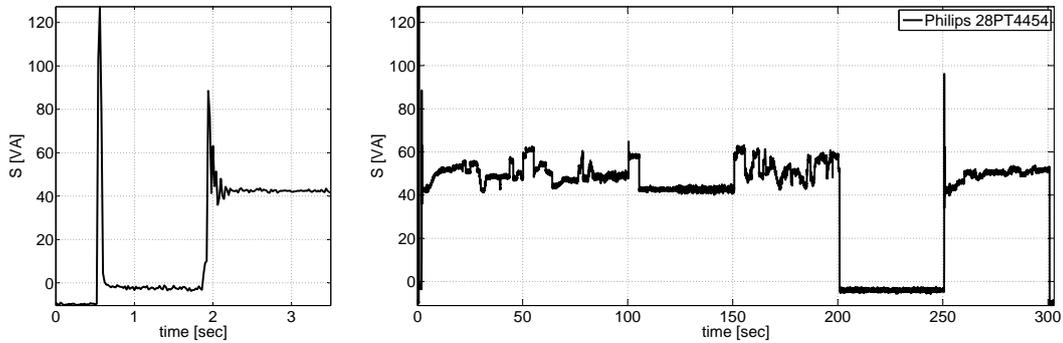


Figure A.6 – Power consumption of a TV. On the left side the transient response is shown and on the right side a typical scenario of the usage of a TV is shown. From 200 s until 250 s the TV was turned in stand-by mode.

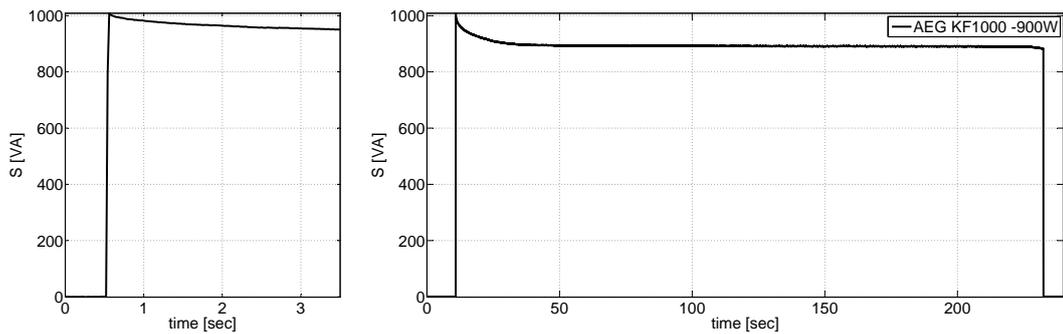


Figure A.7 – Power consumption of a coffee machine. On the left side the transient response is shown its a normal on/ off appliance with two states.

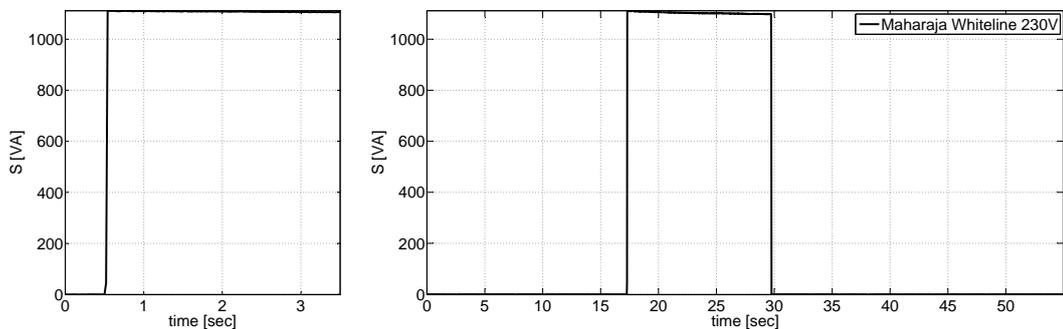


Figure A.8 – Power consumption of a electric iron. It looks similar like the coffee machine. Both appliances has a heating element.

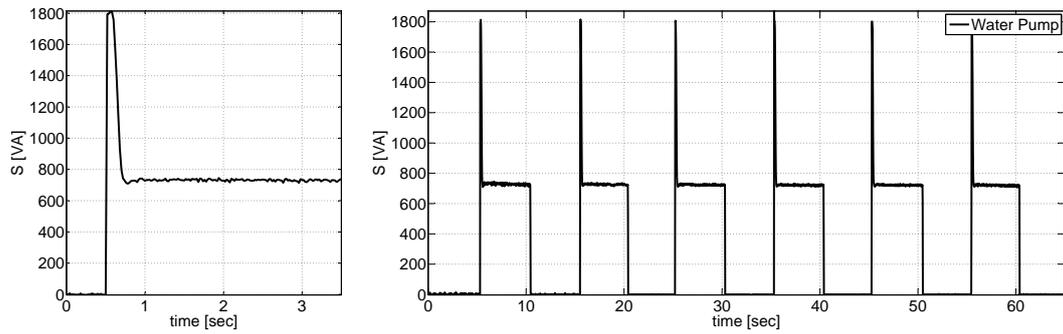


Figure A.9 – Power consumption of a water pump. It has a significant overshoot. This comes from the inductive behavior of this kind of appliance.

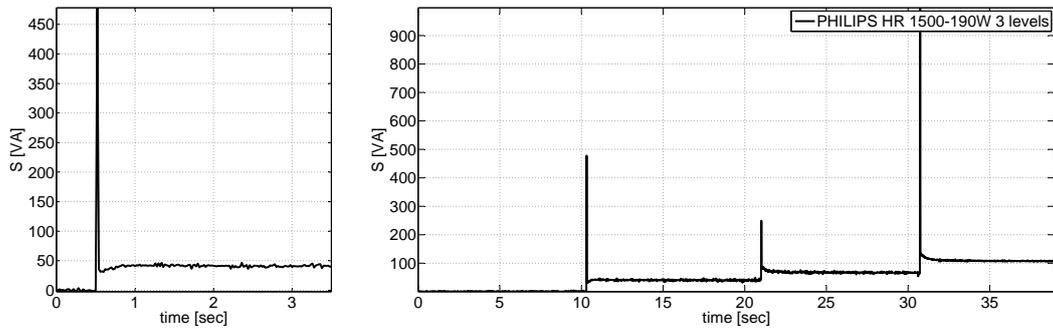


Figure A.10 – Power consumption of a mixer. The mixer has 3 working levels with different power consumption.

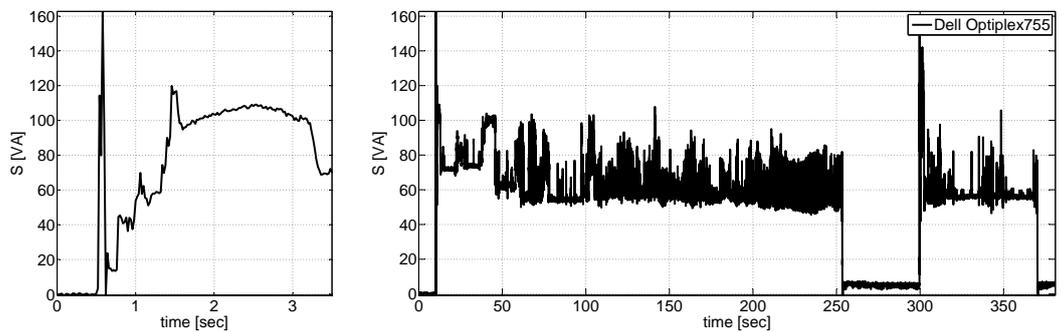


Figure A.11 – Power consumption of a PC. A PC is a more complex appliance. The power consumption depends from different criteria like CPU. From 250s until 300s the PC was turned in stand-by mode.

List of Figures

1.1	Two methods for measuring the energy consumption of electrical appliances in residential buildings.	3
1.2	Overview of the research area NALM at ReSP.	6
1.3	Part of the system for the disaggregation of appliances. It shows the measurement system, the event detector and the classifier. Working on these three blocks is part of that thesis.	7
2.1	Principle of the STFT. The signal is subjected stepwise at the times $t = \tau$ by a windowed FT. For each considered time, the FT provides all frequencies up to half the sampling frequency.	15
2.2	Simulation results of the FT and the STFT for two different switched appliances. In the first row the input signal is shown. In the second row the FT of the input signal is shown and in the last row the results of the STFT is shown.	17
3.1	Overview of Methods for NALM. The methods can be divided into three groups. The traditional ones are the Steady State and the Transient State analyze. In the last years other approaches are developed which deals not directly in these two fields. They can be placed in the third group.	20
3.2	Example of the power consumption of a household. Different appliances were switched on and off.	21
4.1	Structure for the disaggregation of appliances in the load profile of residential buildings. The first block is described. The measurement system measures the entry signals $u(t)$ and $i(t)$. After the calculation of other signals and a pre-conditioning of them, these signals will be stored in a database. Later they can be used for the development of the algorithms for the event detector and the classifier.	29
4.2	Model for the appliances in the load profile of a residential building. The household is part of the external grid which can be modeled as a real voltage source and the resistance of the grid losses.	29
4.3	Effective voltage across the appliances of a real measurement. Based on the voltage divider (Z_S and Z_L) the value of the voltage varies depending on the active appliances.	33
4.4	Modell of a real measurement. The signal $s(t)$ are superimposed by noise $n(t)$. By digitizing a further change is made to the signal.	34
4.5	Three phase test bench system. It is possible to simulate scenarios in the laboratory. It is also possible to record real measurements in residential buildings.	38

4.6	Measurement principle of the measurement boxes for one phase.	39
4.7	Three phase measurement boxes. The signals $u(t)$ and $i(t)$ are conditioned for the AD-conversion. It is possible to measure the signals from all three phases.	40
4.8	Analyzing the measurement system. With the pre-defined input signal $s(t)$ and the measured output signal $y(t)$ the transfer characteristic of the measurement box can be determined.	41
4.9	Bode diagram of the voltage channels of the 1st measurement box . . .	42
4.10	Bode diagram of the current channels of the 1st measurement box. . .	42
4.11	Bode diagram of the voltage channels of the 2nd measurement box. . .	43
4.12	Bode diagram of the current channels of the 2nd measurement box. . .	43
4.13	Switching and detection boxes.	45
4.14	One Phase measurement box. The box can be plugged directly between the appliance and the power outlet.	46
4.15	Example of a measurement with the Fluke and the own developed measurement system.	46
4.16	Pictures of a performed measurement in the laboratory.	49
4.17	Different transient responses of appliances.	51
4.18	Active (P), reactive (Q) and apparent power (S) of different appliances.	53
4.19	Example of switching cycles from on-off appliances. On the left side the power consumption of an automatically switched refrigerator is shown. On the right side the power consumption of a manually switched lamp is shown.	53
4.20	Cycle-operating of a washing machine. It can be modeled as a combination of a water-heater, motor and control electronic.	54
4.21	Cycle-operating of a dishwasher. It can be modeled as a combination of a water-heater, motor and control electronic.	54
4.22	Simulation of a median filter. The first row presents the input signal. In the second row the median filter outputs are shown.	56
4.23	Simulation of a median filter. The first row presents the input signal. In the second row the median filter outputs are shown.	56
4.24	Filter characteristic of the median filter. Example of a real measurement with a peak as distortion and different switching events of appliances.	57
5.1	Disaggregation system: the event detector.	61
5.2	In NALM two kinds of events exist, switching-on and switching-off events. The figure shows an example of a real measurement. In this case a water-heater, a hair-dryer, a refrigerator and a radio were switched on.	62
5.3	Possible switching events with and without disturbances.	63
5.4	Block diagram of the event detector as part of the disaggregation chain.	69
5.5	Masks of the three high-pass filters. The masks $f_{\Delta_{2n}}$ and $f_{\Delta_{3n}}$ based on the gradient operator. These are point-symmetric. The laplace mask f_{Lap} based on the Laplace Operator. This mask is axis-symmetric. . .	74
5.6	Simulation of the high-pass filter. The first row presents the signal with AWGN. The second row the output of the filter with $f_{\Delta_{2n}}$ mask (5.11). Third row output of the advanced filter with $f_{\Delta_{3n}}$ mask (5.13) and the fourth row present the output of the filter with f_{Lap} mask (5.18). . . .	76

5.7	Output of the filters for different gradients of the events.	77
5.8	Function of the first filter with different standard deviation σ . The mask is the second derivative from a Gaussian function.	79
5.9	Function of the second filter with different standard deviation σ . The mask is the Hilbert transform of the second derivative from a gaussian function.	80
5.10	Block diagramm of the approach from Perona with parameters.	80
5.11	Simulation of the Perona Filter. The standard deviation of the filter is $\sigma = 2$. The first row presents the signal with AWGN. In the second and third row the outputs of the two filters are shown (cf. (5.26), (5.27)). The last row shows the output of the decision function (cf. (5.25)). . .	81
5.12	Output of the filters for different gradients of the events.	82
5.13	Simulation of the Perona Filter with different length of the filter, depended from the standard deviation σ of the gaussian function. Here an idealized signal was used, were the $SNR \rightarrow \infty$	83
5.14	Simulation of the Perona Filter with different length of the filter, depended from the standard deviation σ of the Gaussian function. Here an idealized signal was used, were the $SNR = 1,43$	83
5.15	Simulation result for the STFT with a Hamming window of size 10. The input signal is idealized without noise.	85
5.16	Simulation result for the STFT with a Hamming window of size 40. The input signal is idealized without noise.	86
5.17	Simulation result for the STFT with a Hamming window of size 10. The signal was superimposed by noise. It is shown, that the events can be detected. But the influence of noise is increasing and FP are detected.	86
5.18	Simulation result for the STFT with a Hamming window of size 40. The signal was superimposed by noise. It is shown, that the events are detected.	87
5.19	Simulation results of the STFT for different gradients of the switching on event. On the left side the results for a gradient with $\Delta = \frac{5}{1}$ is shown. In the middle the gradient is $\Delta = \frac{5}{2}$ and on the right side the gradient is $\Delta = \frac{5}{3}$	88
5.20	Input and output signal of the decision function. For the mask of the high pass filter the $f_{\Delta_{3n}}$ mask is used. The input signal has 3 amplitude values $A_1 = 5$ W, $A_2 = 10$ W and $A_3 = 15$ W. With a standard deviation of $\sigma = 1,778$ the SNRs for the different amplitudes are $SNR_{A_1} = 1,48$, $SNR_{A_2} = 4,49$ and $SNR_{A_3} = 6,25$	89
5.21	Simulation results of the classifier function. In the first row the TPR in dependence from the threshold ΔP_{th} is shown. In the second row the PPV over the TPR is plotted.	90
5.22	Example of a real measurement. In the first row the power signal Q, P and S are shown. In the second row the output of the event detector using a high pass filter with $f_{\Delta_{3n}}$ mask for the three different input signals is shown.	92

5.23	Outputs of the three approaches for a real measurement signal. In the first row the signal P is shown. In the second row the outputs of the HP Filters, in the third row the output of the Perona Filter and in the last row the output value of the STFT approach is shown.	93
5.24	PPV over TPR for the three approaches to find the optimal parameter values. (a) Filter-Approach with three different masks, (b) Peron-Approach with three different σ values and (c) STFT-Approach with three different window length.	93
6.1	Structure for the disaggregation of appliances in the load profile of residential buildings. Here the third block is described. The classifier assigns each event to a class of appliances.	98
6.2	Model of a biological neuron.	99
6.3	Model of an artificial neuron and its compact structure. This simple form is called perceptron.	101
6.4	Transfer Functions: Sigmoid Function $\alpha = 1, \beta = 0$ (—); Sigmoid Function $\alpha = 2, \beta = 1$ (- -); Sigmoid Function $\alpha = 1, \beta = -1$ (...); Step Function (xxx).	102
6.5	ANN, Feed-Forward-Structure	103
6.6	Multilayer Perceptron	104
6.7	Adaline with Widrow-Hoff learning rule.	107
6.8	Block diagram for the error e calculation.	108
6.9	ADALINE for the harmonics estimation.	110
6.10	Results for the parameter estimation. In the first row the desired signal is shown. In the second row it is shown that the error is decreasing so the algorithm converges. In the last row the convergence of the harmonic values are shown.	111
6.11	Weights of the ADALINE after the learning phase. The weights are corresponds to the harmonics of the desired input signal $p(t)$	111
6.12	Power consumption of a household. During the night phase the switching events of the refrigerator are visible.	112
6.13	Different switching on phases of a refrigerator, with the delta time of the two approaches.	113
6.14	Results for the feature extraction with an ADALINE. With real switching cycles of the refrigerator the true values of the vector \mathbf{w} are calculated.	113
6.15	Example of the results for the classification. Approximately 12 h are presented of the load profile of the measured household.	115
6.16	Diagram of the signal flow from the training, validation and test of the ANN.	116
A.1	Pictures of the measurement system. Using in the laboratory environment to create a big data-set of measurements in a short time.	130
A.2	Measurement in a residential building. The sensors are plugged directly to the household entry point. The voltage and current of three phases were measured.	131
A.3	Instantaneous power of a incandescent lamp and an energy saving lamp.	132
A.4	Time series of the current for one class of appliances.	133
A.5	Harmonics of current signal for one class of appliances.	133

List of Tables

2.1	Contingency table, for the example of the event detector.	12
4.1	SNR of the two boxes	44
4.2	Overview of the important appliances in residential buildings, divided into the type and the switching behavior.	48
4.3	Overview of the important appliances in residential buildings, divided into the type of their power consumption (NRW, 2006).	48
5.1	Results for the event detection methods.	94
6.1	Results for the classification. Two algorithms were analyzed. The results of the approach of Hart are shown in the first line. The results using the ADALINE approach are shown in the second line.	114
6.2	First part of the classification results for the measurements using a MLP. Two features are compared.	117
6.3	TPR of the classification of the appliances with ANN and different input features. Two features are compared.	117
A.1	Overview of the most important signals of a resistance R and an impedance Z for passive sinusoidal current for a two-terminal network.	128
A.2	Overview of the most important signals of an inductance L and a capacitance C for a passive sinusoidal current two-terminal network.	128
A.3	Overview of the current available measurements from the laboratory simulation.	129
A.4	Overview of the current available measurements from the households measurements.	131

B Bibliography

- K. D. Anderson and M. E. Berges.
Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research.
Proceedings of the 2nd Workshop on Data Mining Applications in Sustainability, Beijing, China, 2012.
- K. D. Anderson, M. E. Berges, A. Ocneanu, and D. Benitez.
Event detection for non intrusive load monitoring.
IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, 2012.
- M. Baranski.
Energie-Monitoring im privaten Haushalt: Dissertation.
CUVILLIER VERLAG, Göttingen, 2006.
ISBN 3-86727-020-1.
- M. Basseville and I. V. Nikiforov.
Detection of Abrupt Changes: Theory and Application.
Prentice Hall, 1993.
- D. Benyoucef, T. Bier, and P. Klein.
Planning of energy production and management of energy resources with smart-meters.
Advances in Energy Engineering (ICAEE), International Conference on, Beijing, China, pages 170–173, 2010a.
- D. Benyoucef, P. Klein, and T. Bier.
Smart meter with non-intrusive load monitoring for use in smart homes.
Energy Conference and Exhibition (EnergyCon), IEEE International, Bahrain, pages 96–101, 2010b.
- D. Benyoucef, P. Klein, and T. Bier.
Smart metering: Disaggregation von endverbrauchern.
VDE-Kongress 2010. Smart Cities. Leipzig : VDE VERLAG GMBH Berlin, pages 1–6, 2010c.
- D. Benyoucef, T. Bier, and P. Klein.
Smart metering: Disaggregation von elektrischen verbrauchern im lastgang privaten wohnraums.
Internationaler ETG-Kongresses, ETG-Fachbericht Band, Würzburg, (130), 2011a.
- D. Benyoucef, P. Klein, and T. Bier.
Disaggregation of electrical appliances: Development and verification of detection algorithms.

2. *National Conference on Fractional Order Systems and their Applications. Tizi-Ouzou, Algerien*, 2011b.
- D. Benyoucef, T. Bier, and P. Klein.
Smart meter: Artificial neural network for disaggregation of electrical appliances.
ICPRAM 2012: International Conference on Pattern Recognition Application and Methods. Vilamoura, Portugal, pages 546–550, 2012a.
- D. Benyoucef, T. Bier, and P. Klein.
Test system for disaggregation algorithms for use in smart meters.
Energy Procedia, Elsevier Science, 14:662–667, 2012b.
- P. Bertoldi and B. Atanasiu.
Electricity consumption and efficiency trends in the enlarged european union.
Institute for Environment and Sustainability, JRC European Commission, 2006.
- T. Bier.
Smart metering: Datenerhebung und merkmalsextraktion.
Master Thesis, University Furtwangen, 2010.
- T. Bier, D. Ould-Abdeslam, J. Merckle, and D. Benyoucef.
Smart meter systems detection & classification using artificial neural networks.
IEEE Proceedings of the IECON, Montreal, Canada, 2012.
- T. Bier, D. Benyoucef, D. Ould-Abdeslam, J. Merckle, and P. Klein.
Smart meter systems measurements for the verification of the detection & classification algorithms.
IEEE Proceedings of the IECON, Vienna, Austria, 2013.
- C. Bishop.
Pattern Recognition and Machine Learning.
Springer, 2006.
ISBN 0-387-31073-8.
- Bundesministeriums der Justiz, editor.
Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz - EnWG), 2005.
- J. Canny.
A computational approach to edge detection.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, 8(6):679–698, 1986.
ISSN 0162-8828.
- A. Cichocki and R. Unbehauen.
Neural networks for optimization and signal processing.
Stuttgart : Teubner, Chichester et al.: Wiley, 1993.
- R. Cox, S. Leeb, S. Shaw, and L. Norford.
Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion.
Applied Power Electronics Conference and Exposition, APEC '06. Twenty-First Annual IEEE, 2006.

- Dawei He, Weixuan Lin, Nan Liu, R. Harley, and T. Habetler.
Incorporating non-intrusive load monitoring into building level demand response.
Smart Grid, IEEE Transactions on, 4(4):1870–1877, 2013.
- D. Demigny.
On optimal linear filtering for edge detection.
IEEE TRANSACTIONS ON IMAGE PROCESSING, 11(7):728–737, 2002.
- digitalStrom, 2012.
URL <http://www.digitalstrom.com/?gclid=CJTA4ZP4vrMCFYq7zAodciQAvw>.
- K. Ehrhardt-Martinez, K. Donnelly, and J. Laitner.
Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities.
American Council for an Energy-Efficient Economy, Technical report, Research Report E105, 2010.
- L. Farinaccio and R. Zmeureanu.
Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses.
Energy and Buildings, Elsevier Science, 30(3):245–259, 1999.
- R. Fernandes, I. N. da Silva, and M. Oleskovicz.
Load profile identification interface for consumer online monitoring purposes in smart grids.
Industrial Informatics, IEEE Transactions on., 9(3):1507–1517, 2013.
- S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood.
Creative user-centered visualization design for energy analysts and modelers.
Visualization and Computer Graphics, IEEE Transactions on, 19(12):2516–2525, 2013.
- G. Hart.
Prototype nonintrusive appliance load monitor.
MIT Energy Laboratory Technical Report, and Electric Power Research Institute Technical Report, 1985.
- G. Hart.
Nonintrusive appliance load monitoring.
Proceedings of the IEEE, 80(12):1870–1891, 1992.
- T. Hassan, F. Javed, and N. Arshad.
An empirical investigation of v-i trajectory based load signatures for non-intrusive load monitoring.
Smart Grid, IEEE Transactions on, 5(2):870–878, 2014.
- D. M. Hawkins and D. H. Olwell.
Cumulative Sum Charts and Charting for Quality Improvement.
Springer, New York, 1998.
ISBN 0-387-98365-1.

- S. Haykin.
Adaptive filter theory.
Prentice-Hall, Great Britain: Pearson Education International, 2002.
- D. Hebb.
Organisation of Behavior.
Wiley & Sons, New York, 1949.
- J. Hopfield and D. Tank.
Neural computation of decision in optimization problems.
Biological Cybernetics, Springer-Verlag, 52:141–152, 1985.
- International Energy Agency.
Energy statistics manual: Electricity/heat in world, 2008.
URL <http://www.iea.org/>.
- Y. Jimenez, C. Duarte, J. Petit, and G. Carrillo.
Feature extraction for nonintrusive load monitoring based on s-transform.
Power Systems Conference (PSC), Clemson University, pages 1–5, 2014.
- D. Jung, A. Savvides, and A. Bamis.
Tracking appliance usage information in residential settings using off-the-shelf low-frequency meters.
Design Automation Conference (DAC), 49th ACM/EDAC/IEEE, pages 163–168, 2012.
- K.-D. Kammeyer and K. Kroschel.
Digitale Signalverarbeitung: Filterung und Spektralanalyse mit MATLAB-Übungen.
Teubner, 8 edition, 2012.
- S. Kazakidis, A. Kokkosis, K. Moustris, and A. G. Paliatsos.
Electricity consumption prognosis with the combination of smart metering and artificial neural networks.
Power Generation, Transmission, Distribution and Energy Conversion (MED-POWER 2012), 8th Mediterranean Conference on., pages 1–6, 2012.
- P. Klein.
Smart metering: Entwicklung einer disaggregationsstrategie.
Master Thesis, University Furtwangen, 2011.
- P. Klein, J. Merckle, D. Benyoucef, and T. Bier.
Test bench and quality measures for nilm algorithms.
IEEE 39th Annual Conference of the IEEE Industrial Electronics Society, Vienna, Austria, 2013.
- J. Kolter and M. Johnson.
Redd: A public data set for energy disaggregation research.
SustKDD, San Diego, CA, USA, 2011.
- K. Kroschel, G. Rigoll, and B. Schuller.
Statistische Informationstechnik.
Springer, 2011.

- H. Lam, W. Lee, G. Fung, F. Chan, and M. Lucente.
Exploration on load signatures.
International Conference on Electrical Engineering (ICEE), 2004.
- H. Lam, G. Fung, and W. Lee.
A novel method to construct taxonomy of electrical appliances based on load signatures.
IEEE, supported by the postgraduate studentship of the University of Hong Kong, and CLP Research Institute Ltd., 2007.
- C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong.
Power signature analysis.
Power and Energy Magazine, IEEE, 1(2), 2003.
- K. Lee, L. Norford, and S. Leeb.
Development of a functioning centrally located electrical-load monitor.
Technical Report, Massachusetts Institute of Technology, 2003.
- S. Leeb, S. Shaw, and J. Kirtley.
Transient event detection in spectral envelope estimates for nonintrusive load monitoring.
Power Delivery, IEEE Transactions on, 10(3), 1995.
- Y.-H. Lin and M.-S. Tsai.
Development of an improved time–frequency analysis-based nonintrusive load monitor for load demand identification.
Instrumentation and Measurement, IEEE Transaction on, 63(6):1470–1483, 2014.
- S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic.
Ampds: A public dataset for load disaggregation and eco-feedback research.
Electrical Power & Energy Conference (EPEC), IEEE, pages 1–6, 2013.
- M. Marceau and R. Zmeureanu.
Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings.
Energy Conversion & Management, Elsevier Science, 41(3):1389–1403, 2000.
- H. Moon, R. Chellappa, and A. Rosenfeld.
Optimal edge-based shape detection.
Image Processing, IEEE Transactions on, 11(11):1209–1227, 2002.
ISSN 1057-7149.
- A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt.
A robust heart sounds segmentation module based on s-transform.
Biomedical Signal Processing and Control, Elsevier Science, 8(3):273–281, 2013.
- H. Najmeddine, K. Khamlichi Drissi, C. Pasquier, C. Faure, K. Kerroum, A. Diop, T. Jouannet, and M. Michou.
State of art on load monitoring methods.
2nd International Conference on Power and Energy (PECon 08), IEEE, Johor Baharu, Malaysia, 2008.

- Y. Nakano.
Non-intrusive load monitoring system – part 5: Performance test at real households.
System Engineering Research Laboratory, 2005.
- NRW.
Mean electricity costs averaged over all household sizes.
Press Info from, 2006.
URL http://www.ea-nrw.de/_infopool/page.asp?InfoID=4106.
- T. Onoda, G. Rätsch, and K.-R. Müller.
Applying support vector machines and boosting to a non-intrusive monitoring system for household electric appliances with inverters.
2000.
URL <http://citeseer.ist.psu.edu/261408.html>.
- F. Paradiso, F. Paganelli, A. Luchetta, D. Giuli, and P. Castrogiovanni.
Ann-based appliance recognition from low-frequency energy monitoring data.
World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE 14th International Symposium and Workshops on a, pages 1–6, 2013.
- S. Pattem.
Unsupervised disaggregation for non-intrusive load monitoring.
Machine Learning and Applications (ICMLA), 2012 11th International Conference on., 2:515–520, 2012.
- P. Perona and J. Malik.
Detecting and localizing edges composed of steps, peaks and roofs.
Computer Vision, 1990. Proceedings, Third International Conference on., pages 52–57, 1990.
- H. Pihala.
Non-intrusive appliance load monitoring sytem based on modern kwh-meter.
Technical Research Center of Finland, 1998.
- J. Powers, B. Margossian, and B. Smith.
Using a rule-based algorithm to disaggregate end-use load profiles from premise-level data.
Computer Applications in Power, IEEE, 4(2):42–47, 1991.
- L. G. Roberts.
Machine perception of three-dimensional solids.
Thesis (Ph. D.), Massachusetts Institute of Technology, 1963.
- F. Rosenblatt.
The perceptron. a probabilistic model for information storage and organization in the brain.
Psychological Review, 65(6):386–408, 1958.
- A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O’Hare.
Real-time recognition and profiling of appliances through a single electricity sensor.
Sensor Mesh and Ad Hoc Communications and Networks (SECON), 7th Annual IEEE Communications Society Conference on, pages 1–9, 2010.

- S. Shaw.
System identification and modeling for nonintrusive load diagnostics.
Thesis (Ph.D.), Massachusetts Institute of Technology, 2000.
- Y. Sowa, A. D. Rowe, M. C. Leake, T. Yakushi, M. Homma, A. Ishijima, and R. M. Berry.
Direct observation of steps in rotation of the bacterial flagellar motor.
Nature, 437(7060):916–919, 2005.
doi: 10.1038/nature04003.
URL <http://dx.doi.org/10.1038/nature04003>.
- R. Stockwell, L. Mansinha, and R. Lowe.
Localization of the complex spectrum: the s transform.
Signal Processing, IEEE Transactions on., 44(4):998–1001, 1996.
- VDE, editor.
Smart Energy 2020, 2010. VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V.
- S. Vogts.
Smart metering: Entwicklung einer disaggregationsstrategie.
Master Thesis, University Furtwangen, 2011.
- W. Wichakool, Z. Remscrim, U. Orji, and S. Leeb.
Smart metering of variable power loads,.
Smart Grid, IEEE Transactions on, PP(99):1–10, 2014.
- B. Widrow and M. Hoff.
Adaptive switching circuits.
IRE WESCON Convention Record, (4):96–104, 1960.
- W. Xu and M. Dong.
Tracking energy consumptions of home appliances using electrical signature data.
Power and Energy Society General Meeting, 2012 IEEE, pages 1–5, 2012.
- M. Zeifman and K. Roth.
Nonintrusive appliance load monitoring: Review and outlook.
Consumer Electronics, IEEE Transactions on., 57(1):76–84, 2011.